

Contents

Research Articles

Orli Dahan: <i>There IS a Question of Physicalism</i>	542
Matej Drobnák: <i>Do We Share a Language? Communitarism and Its Challenges</i>	572
Bartosz Kaluziński: <i>Genuinely Constitutive Rules</i>	597
Adam Greif: <i>The Morality of Euthanasia</i>	612
David Černín: <i>Historical Antirealism and the Past as a Fictional Model</i>	635

There IS a Question of Physicalism


Orli Dahan*

Received: 12 June 2018 / Accepted: 30 November 2018

Abstract: The most common catchphrase of physicalism is: “everything is physical”. According to Hempel’s Dilemma, however, physicalism is an ill-formed thesis because it can offer no account of the physics to which it refers: current physics will definitely be revised in the future, and we do not yet know the nature of future physics. The dilemma arises due to our difficulty to set the boundaries of the concept ‘physical.’ In order to confront the dilemma, a physicalist must ensure that physics is not going to broaden itself artificially (or in some trivial way) to become complete—perhaps by adding non-reductive mental entities to elementary physical theory, making it impossible to distinguish physicalism from dualism. I offer a solution to the dilemma which is a version of the ‘via negativa’ (standardly taken to be a stipulation that the physical not include the mental), albeit one that is specified and worked out in a distinctive way. My suggested formulation of the physicalist hypothesis allows us to establish a refutation condition of physicalism. The refutation condition is general and not only dualistic. Consequently, the physicalist can choose the second horn of the dilemma, and hold that physicalism is indeed refutable (and not a trivial thesis).

Keywords: Completeness of physics; consciousness; dualism; Hempel’s Dilemma; physicalism; via negativa.

* Tel-Hai College

 Faculty of Humanities and Social Sciences, Tel-Hai College, Upper Galilee, 12208, Israel

 orlydah@telhai.ac.il



1. Introduction

One primary claim of physicalism is that physical theory guides us in answering the question “what does the world contain?” because, according to physicalism, everything is physical. Stated this way, the thesis of physicalism faces a dilemma, referred to in current literature as ‘Hempel’s Dilemma.’ The dilemma, which deals with the meaning of the concept ‘physical,’ arises because on the one hand contemporary physical theory is incomplete, but on the other hand we know nothing about the character and properties of the future and supposedly complete physical theory. In other words, the question that a physicalist must ask herself, according to the dilemma, is this: which *exact* physical theory guides the physicalist’s ontology? (Hempel 1966; Crane and Mellor 1990).

According to Hempel’s Dilemma, there are two ways to interpret the primary claim of physicalism, and both are troublesome:

1. Physicalism is the thesis that the world contains *only* entities assumed and defined by contemporary physics, and that these entities behave according to the laws of *contemporary* physical theory.
2. Physicalism is the thesis that the world contains *only* entities that will be assumed and defined by future physics, and that these entities will behave according to the laws of *future-and-complete* physical theory.

The first horn of the dilemma leads to a false thesis because contemporary physics is incomplete; at this point, physical theories do not enable us to provide a complete explanation of all that exists in the world. However, the second horn is also problematic because it ostensibly leads to a meaningless, irrefutable, or trivial thesis: we do not really know what we are committed to as physicalists because we do not know enough about future-and-complete physics. The advocate of Hempel’s Dilemma is concerned that to become complete, physical science might *broaden* its scope in different ways, perhaps by adding non-reductive mental entities to elementary physical theory and, consequently, we might not be able to distinguish physicalism from dualism.

Discussions regarding Hempel’s Dilemma are common in contemporary literature on physicalism and the mind-body problem [see for example

(Bokulich 2011); (Crook and Gillett 2001); (Dowell 2006); (Fiorese 2015); (Gillett and Witmer 2001); (Montero 2001); (Ney 2008); (Prelević 2017); (Prelević 2018); (Stoljar 2010); (Wilson 2006); (Worley 2006)]. The responses to the dilemma can be divided as follows:¹

1. *“Currentism”*: Taking the first horn of the dilemma while dismissing the claim that it makes physicalism a false thesis: some physicalists think that although contemporary physics is incomplete, it is reasonable to assume that its main characteristics will remain more or less the same. For example, Lewis (1983/1999, 33–34) argues that the physical theory that guides ontology is an improved version of contemporary physics that is not exceptionally different from contemporary physics. Bokulich (2011) argues that our knowledge of current physics is sufficient for offering a physicalist ontology of the mind. According to Bokulich, we have solid scientific evidence that future physics will be irrelevant to the mind-body problem because mental processes are part of the well-understood domains of applicability of current physical theories.² Vicente (2011) proposes to construe current physics minimally according to the following assertions: some properties are conserved quantities, those quantities are possessed by bodies, and their distribution and exchange are mediated by forces. According to Vicente, the construal of current physics allows for an adequate definition of the physical with regard to Hempel’s Dilemma.
2. *“Futurism”*: Taking the second horn of the dilemma while inserting a constraint on the formulation of physicalism: a popular suggestion is to include in the formulation of physicalism the constraint

¹ Prelević (2017; 2018) divides the strategies of dealing with Hempel’s Dilemma into three: defending currentism, defending futurism or trying to avoid the dilemma by claiming that physicalism is not a thesis. I agree with this general division but add a few more options for rejecting or avoiding the dilemma.

² Bokulich’s approach can be ascribed to a second option: that of choosing the first horn of the dilemma *only about the mind-body problem*. However, it is worth mentioning that physicalism is a general thesis, which is not limited to the mind-body problem, even if that is the current focus in the literature.

“without fundamental mental properties in future-and-complete physics”. This solution ensures the possibility of distinguishing physicalism from dualism in the future [see for example (Wilson 2006); (Worley 2006)]. According to Ney (2008), if non-reductive mental entities are included in future-and-complete physical theory, then the dualist will know that she had been right all along, and the physicalist will realize she had been wrong. Therefore, taking the second horn of the dilemma does not make physicalism irrefutable, despite the claim of the dilemma.³

3. *Avoiding the dilemma by defining the physical as non-mental ('via negativa')*: This solution proposes to render the term ‘mental’ as fundamental and to characterize the physical as ‘non-mental,’ i.e. defining the term ‘physical’ in a negative way [for more detail see (Crook and Gillett 2001); (Gillett and Witmer 2001); (Montero 2001); (Montero and Papineau 2005); (Wilson 2006); (Worley 2006)]. This suggestion has much in common with the previous one, and, indeed, the differences between the two solutions are rather subtle. The *via negativa* solution can be viewed as *circumventing* the dilemma because it avoids the need to link physicalism to any particular physical theory. Hence, the dilemma becomes irrelevant. Fiorese (2015) argues in favor of the ‘*via negativa*,’ holding that either the *via negativa* is valid, or there is, indeed, no version of physicalism deserving of the name.⁴ Moreover, according to Prelević (2017), although this view was originally introduced as a version of futurism, it can also be incorporated into the hard-core of the physicalist research programme.
4. *Avoiding the dilemma by rejecting the claim that physicalism is a thesis*: according to Prelević (2017), there are two strategies of avoiding the dilemma by arguing that physicalism is actually not a thesis: the

³ Ney (2008) also argues that physicalism could be better seen as an attitude instead of an ontological thesis and, in this way, to avoid the problems derived from Hempel’s Dilemma. Ney’s suggested attitude is based upon a commitment to construe one’s ontology according to what physics says exists [for a detailed discussion about this view see (Prelević 2017)].

⁴ For a rather recent criticism of the so-called ‘*via negativa*,’ see (Vicente 2011).

attitudinal approach, according to which physicalism is a stance or an attitude, and the Lakatosian approach, according to which physicalism is best understood as a research programme. If physicalism is not a thesis, it cannot be true, trivial or empty.

5. *Rejecting the dilemma*: there are several strategies to reject the dilemma while still holding that physicalism is a thesis. For example, Dowell (2006) explains what makes a theory a physical one in terms of the hallmarks of scientific theories, and suggests tying physicalism's ontological commitments to our best methods for justifying our beliefs about the natural world. This solution rules out some entities as falling within the extension of 'the physical' and thus gives 'physicalism' more content than made apparent by discussions of the second horn of Hempel's Dilemma.⁵ Stoljar (2010) argues that the dilemma collapses because it has a third premise (which is almost always overlooked in the literature). The third premise encourages us to choose between the two horns of the dilemma. According to Stoljar, this third premise is mistakenly believed to be a logical truth, while in fact it is not a logical truth, but a substantive falsehood (Stoljar 2010, 105–6).⁶

Each of the proffered solutions is based upon the same basic premise: that empirical physical science will guide us in answering the ontological question of what is in the world. However, I believe that most advocates of Hempel's Dilemma would not resolve the dilemma by reconciling the two horns, as this can be seen as dodging. Moreover, most advocates of Hempel's Dilemma also would not choose the first horn of the dilemma, no

⁵ Dowell's view (Dowell 2006) can also be counted as a version of futurism. But I have defined futurism as taking the second horn of the dilemma while inserting a constraint on the formulation of physicalism. Dowell rejects inserting this constraint and, in fact, allows that postulating consciousness at the fundamental level would be in accordance with physicalism. Dowell's view will be discussed further in Section 7.

⁶ Stoljar is arguing this from some metaphysical considerations about a twin-physical world (Stoljar 2010, 77). In this sense, his view is different from the view of Dowell (2006), which is more empirically oriented. However, the difference in motivation does not change the result: both views reject the dilemma.

matter how convincing the idea that current physics will not change significantly, particularly concerning the mind. Quite justifiably, most philosophers will remain unconvinced in the face of the uncertainty of something empirical like physics. At the end of the nineteenth century, it seemed that humankind was close to resolving physical theory only to witness its radical change.

In a way, defining physicalism not as a thesis, but as an attitude or a research programme has many advantages [for a detailed discussion see (Prelević 2017, 2018)]. However, doing so weakens the core idea of physicalism because there is no doubt that physicalism is generally understood as a thesis (or a hypothesis) regarding the actual world. Thus, in this paper, I offer a solution to the dilemma that helps physicalism maintain the status of a thesis.

I believe that a more precise solution, related to the *via negativa* strategy, exists in the empirical spirit of the solutions mentioned above. To answer the advocate of Hempel's Dilemma, I propose a clarification of the physicalist's primary hypothesis in the context of the mind-body problem and in two other contexts as well (vitalism and emergence). First, in the next section, I briefly discuss physics' pretense to completeness, which further justifies the formulation I am about to suggest, together with its putative implications. In Section 3, I introduce my proposal. In Section 4, I establish a refutation condition of physicalism and argue that it allows us to choose the second horn of the dilemma without making physicalism irrefutable. In Section 5, I show that the refutation condition I offer can disarm the motivation of the advocate of Hempel's Dilemma. In Section 6, I sketch briefly the putative implications of my solution. In Section 7, I emphasize why the way I confront the dilemma is a version of the *via negativa*, albeit one that is specified and worked out in a distinctive way. In Section 8, I offer concluding remarks.

2. A note about physics' pretense to completeness

A remark is needed regarding physics' *pretense* to completeness. First, because it seems that this is the most successful argument in favor of physicalism (Papineau 2001). Second, because this is the reason for designing

the physicalist's hypothesis and its refutation condition that I intend to propose.

With regard to the completeness of physics, the point can be phrased in this way: we assume that in order to explain the behavior of billiard balls, tables, hurricanes, and any other parts of the world (objects and phenomena) that do not involve mental aspects—we do not need mental laws (in addition to some complete physical theory). We also assume that the same condition applies in the context of explaining mental states; we do not need additional laws, over and above the laws of complete physical theory, to explain mental phenomena.

Of course, it is another question why (or whether) physics is required (in principle) to explain *all there is*. What is the source of this universally explanatory motivation? Can we not assume that physics explains some aspects of reality (say—objects like atoms, quarks, tables, and chairs) while accepting that physics (in principle) cannot explain some aspects of other objects (say—objects with a mental aspect, complex objects like organic cells, communities, and such)? This issue is related to 'the causal closure principle of the physical-domain' (CCP), meaning that every physical event has a physical cause [see (Kim 2006); (Stoljar 2017)]. And so, if a behavior of a man (say reaching for a glass of water and drinking) is physical, then it has only physical causes, no matter how strong our intuitions tell us that this behavior also had mental causes (the "will" to drink, or the "feeling" of thirst).⁷

However, the justification for believing in CCP is not obvious. Is it a metaphysical principle (dealing with causality and determination), or is it an empirical hypothesis based on the laws of physics? While according to Papineau (2001) the belief in CCP is historically related to conservation laws,⁸ Montero (2006) rejects that conservation laws ground the CCP, and

⁷ A non-eliminativist physicalist will insist that the behavior did indeed have mental causes because the mental is physical. Of course, there are no causes in addition to the physical causes.

⁸ In physics, a conservation law states that a particular measurable property of an isolated physical system does not change as the system evolves. Laws that have been confirmed so far (i.e. never been violated) include conservation of mass-energy, conservation of linear momentum, conservation of angular momentum and conservation

Vicente (2006) tries to explain how the CCP could be made to follow from conservation laws using additional premises. On the other hand, Bishop (2006) argues that there is a hidden premise that must be added to the CCP for the causal argument to be sound. The hidden premise is as follows: ‘the only efficacious states and causes are physical ones.’ But since it is indistinguishable from the conclusion of the CCP—the argument begs the question regarding physicalism.

Nevertheless, the pretense of physics to completeness can be phrased differently. For example, if the “Theory of Everything” (the physical theory that would account for everything that there is) has been given to us (in some miraculous way), then the claim of the physicalist is: there is no extra fact in the world that is not derived from that theory.⁹ Kripke phrases physicalism’s claim in a similar way, though much more figuratively. According to Kripke, if physicalism is true, then when God created the world, God had only to fix the elementary particles and set the relationship between them, and everything else occurred automatically (Kripke 1980, 153–54).

In this context, we can see the tight connection between the ideal physical theory, physics’ pretense to completeness, and the metaphysical thesis of physicalism. The physicalist thinks of mental states, in particular, and all other states, in general, as included in the complete physical theory about the world. Certainly, the thesis of physicalism regarding the mind-body problem is not yet confirmed, because it is only a hypothesis:¹⁰ at the moment we do not possess a psycho-physical theory. However, even if physicalism is a hypothesis, it still might be a highly-confirmed hypothesis. Because for physicalists who accept the causal closure argument as sound—it

of electric charge. Additional conservation laws have been confirmed empirically; however, the subject goes beyond the scope of this paper.

⁹ See for example (Redhead 1996, 63–66) on the “theory of everything” as the Holy Grail of modern theoretical physics. The unification programme of fundamental physics will be discussed further in Section 4.

¹⁰ One can ask whether this hypothesis is metaphysical or empirical. Typically, physicalism is a metaphysical view of the nature of reality (or at least this is the way it is discussed within the mind-body literature). However, others think of physicalism more like an empirical hypothesis regarding the actual world [see (Spurrett 2017)].

seems to provide good confirmation of physicalism even in the absence of a psycho-physical theory.

I want to emphasize that in this paper I have no intention of arguing in favor of physicalism. The aim of this paper is only to show that there is, in fact, a question of physicalism—thus the thesis of physicalism is not trivial. I argue that in the future, the thesis of physicalism might be discovered to be false. For this reason, I am not going to justify physics' pretense to completeness or the CCP; for these subjects go beyond the scope of this paper and are indeed in controversy. However, in the context of the completeness of physics, it is worth noting that even in the many diverse branches of physics itself, were we to assume a successful reduction, the reduction would not necessarily be complete. One such branch is thermodynamics: we still cannot derive all thermodynamics laws and phenomena from classical statistical mechanics [for further details see (Hemmo and Shenker 2012)].

And so, there are many arguments and discussions in the literature of the mind-body problem against the thesis of physicalism. Moreover, there are many other issues related to the field of philosophy of science in general that question whether physics can reach completion or can, in general, explain all there is (opponents include strong emergentists). However, the strength of Hempel's Dilemma is its second horn, that physicalism is an empty doctrine and not refutable under any future conditions and circumstances. This claim, I will show, is wrong.

3. The hypothesis of physicalism elucidated in the context of Hempel's Dilemma

I will introduce my formulation of the hypothesis in the context of the mind-body problem while noting that physicalism is a general thesis and the hypothesis can be phrased in other contexts as well (for example, in the context of the relationship between physics and biology or physics and aesthetics). Of course, the consequences of Hempel's Dilemma reach beyond the philosophy of the mind; if the dilemma is sound, it potentially invalidates physicalism as a thesis that is false or has no content at all. A crucial note is that this formulation by itself is not an answer to the dilemma, but

only the first step in the way to construct a refutation condition of physicalism (in the next section).

The hypothesis of physicalism from the mind-body point of view, based on the supervenience thesis,¹¹ emerges out of two premises:

1. *Premise I: regarding the world*—the macroscopic objects in the actual world can be divided into two sets: set a includes things *without* mental properties (for example tables, chairs, and even complex systems and phenomena, such as hurricanes or cells, as long as they do not possess any mental states) and set B includes things *with* mental properties (for example humans). These two sets exhaust the macroscopic objects that exist in the actual world.
2. *Premise II: regarding the future-and-complete physical theory*—the laws of future-and-complete physics will be identical for both objects of set A and of set B.
3. *Therefore, physicalism's primary hypothesis is as follows:* the physical laws of a will (in principle) be sufficient in order to serve as a foundation for building psychological theories. Meaning: the physical laws that apply to objects in set a *are sufficient* to describe and explain the behavior and properties of objects in set B.^{12,13}

¹¹ Supervenience is a relation that is used to describe cases where the upper-level properties of a system are determined by its lower level properties. Most philosophers think of physicalism as a metaphysical thesis, and so it is usually understood to mean that everything logically (or metaphysically) supervenes upon the physical. Others believe the supervenience thesis with regard to physicalism is contingent because physicalism is simply an empirical hypothesis about the actual world. However, this observation is not crucial regarding my argument in this paper.

¹² Also, they will be sufficient for building the foundation of any other scientific theory, for example, thermodynamics, biology, geology, etc.

¹³ This formulation is, in fact, compatible with epiphenomenalism. If mental properties are epiphenomenal, then it will be possible to describe and explain the behavior of objects in set B using physical laws, even though there will be some properties that are not physical. But there are a few reasons to think that epiphenomenalism is not an option for the actual world. For example, Papineau (2016) argues that epiphenomenalism is not an attractive position, for it presents a very

This formulation can be seen as a ‘via negativa’ way to characterize the physical domain; however, I emphasize that this is not the only way of dividing the macroscopic world in an exhaustive manner. One can use other sets of objects to divide the world in an exhaustive manner. Given the context of the mind-body problem and physicalism, I present the first premise this way. From of a biological point of view, I would divide the objects in the world differently.

In fact, I propose to handle the advocate of Hempel’s Dilemma by relating the physical to the macroscopic *objects and properties* that physics is bound to account for, and not by rendering the term ‘mental’ as fundamental and by characterizing the physical as ‘non-mental.’ The recent tendency to focus on Hempel’s Dilemma as an issue for physicalism about the mind neglects the importance of vitalism and emergence for the history of physicalism as well as the thinking about how to distinguish the physical from the non-physical. At times, this focus makes the topic of physicalism seem needlessly parochial or narrow. For this reason, I suggest two additional ways of formulating the physicalist’s hypothesis: a vitalist and a strong emergence formulation.

The hypothesis of physicalism from a biological point of view, based on the supervenience thesis, emerges out of two premises:

odd kind of causal structure: “nature displays no other examples of such one-way causal intercourse between realms.” Moreover, there is an epistemological problem with epiphenomenalism: we cannot explain our knowledge *about* mental states if mental states are just epiphenomenal and have no effect on the world (Braddon-Mitchell and Jackson 1996, 6–7). And, if mentality is not a basic structure of the world, but an evolutionary product (or by-product), it seems hard to explain from an evolutionary point of view the existence of properties that have no effect on us (Braddon-Mitchell and Jackson 1996, 6–7). Given these considerations, it seems reasonable to exclude the epiphenomenalism possibility about the actual world, or even just to prefer physicalism over epiphenomenalism. I am aware of the fact the one cannot be comfortable with an account that allows for the possibility of there being epiphenomenal non-physical properties. However, one should acknowledge that the possibility that something non-physical obeys physical laws is remote or directly nonsensical. For more detail about epiphenomenalism see (Robinson 2015).

1. *Premise I': regarding the world*—the macroscopic objects in the actual world can be divided into two sets: set A' includes things *without* organic properties (for example tables, chairs, and even complex systems and phenomena, such as hurricanes, as long as they do not possess any organic properties) and set B' includes things *with* organic properties (for example humans, amoebas, and bacteria). These two sets exhaust the macroscopic objects that exist in the actual world.
2. *Premise II': regarding the future-and-complete physical theory*—the laws of future-and-complete physics will be identical for both objects of set A' and of set B'.
3. *Therefore, physicalism's primary hypothesis is as follows*: the physical laws of A' will (in principle) be sufficient in order to serve as a foundation for building biological theories. Meaning: the physical laws that apply to objects in set A' *are sufficient* to describe and explain the behavior and properties of the organic objects in set B'.¹⁴

The hypothesis of physicalism from an emergentist point of view,¹⁵ based on the supervenience thesis, emerges out of two premises:

¹⁴ This version of physicalism can be seen as similar to a venerable proposal by Meehl and Sellars (1956) that urged distinguishing a very broad and a more restrictive version of the 'physical' (the '1' and '2' are sub-scripts in the original): physical₁—terms employed in a coherent and adequate descriptive, explanatory account of the spatiotemporal order; physical₂—terms used in the formulation of principles which suffice in principle for the explanation and prediction of inorganic processes. Meehl and Sellars (1956) were also concerned with emergence and wanted to reject an argument that the doctrine is trivially false, arguing instead that it is coherent (and empirically false).

¹⁵ According to Chalmers (2006), the term 'emergence' is used to express at least two different concepts: weak emergence and strong emergence. Weak emergent properties are (in principle) deducible from the low-level properties (perhaps in conjunction with knowledge of initial conditions) while strongly emergent phenomena are systematically determined by low-level facts without being deducible from those facts. Thus, strong emergence has much more radical consequences than weak emergence: "Strong emergence, if it exists, can be used to reject the physicalist picture of the world as fundamentally incomplete. By contrast, weak emergence can be used to support the physicalist picture of the world by showing how all sorts of phenomena

1. *Premise I'*: regarding the world—the macroscopic objects in the actual world can be divided into two sets: set A' includes things *without* emergent properties (for example tables, chairs, and all objects that do not introduce emergent properties)¹⁶ and set B' includes things with emergent properties (for example hurricanes, humans, traffic jams, and communities). These two sets exhaust the macroscopic objects that exist in the actual world.
2. *Premise II'*: regarding the future-and-complete physical theory—the laws of future-and-complete physics will be identical for both objects of set A' and of set B'.
3. *Therefore, physicalism's primary hypothesis is as follows*: the physical laws of A' will (in principle) be sufficient in order to serve as a foundation for building high level theories about emergent objects and phenomena in the world. Meaning: the physical laws that apply to objects in set A' *are sufficient* to describe and explain the behavior and properties of the emergent objects and properties in set B'.

4. Establishing a refutation condition of physicalism and choosing the second horn of the dilemma

According to the aforementioned formulation of the physicalist's claim, I will now explain why a physicalist can choose the second horn of Hempel's

that might seem novel and irreducible at first sight can nevertheless be grounded in underlying simple laws."

¹⁶ It is worth noting that according to some views, tables and chairs can also be viewed as objects that introduce emergent properties [see for example (Teller 1992)]: the macroscopic properties of a table is much different from its microscopic properties. If we take the "naked" emergentist intuition to be that an emergent property of a whole somehow "transcends" the property of the parts—nearly all macroscopic scale objects would be "emergent". If this is the view in our context, then the hypothesis of the physicalist can be formulated slightly different, and divide the world into microscopic objects and macroscopic objects. However, I chose to construct the hypothesis as follows because most strong emergentist's views would not consider tables as emergent in an interesting way [see also (Chalmers 2006)].

Dilemma while rejecting the claim that it makes physicalism irrefutable. When looking at premise II, it becomes clear that if in the future premise II would be refuted, then physicalism would be refuted as well. The physicalist claims that the laws of physics govern the bodies and behavior of humans, and also govern the bodies and behavior of objects such as tables, chairs, billiard balls, and even more complex phenomena such as hurricanes or cells. This is the reason why the parts of the world in which mental states are present (objects of set B) are approximately governed by the laws of contemporary physics. If this hypothesis is to be refuted in the future, for instance, due to a *division of physical laws* between set A and set B, or between set A' and set B', or between set A'' and set B''—then physicalism is refuted. Now, because the dilemma asserts the falsehood or triviality of physicalism, in formulating the thesis as in the previous section, I maintain that the thesis of physicalism is not trivial at all. The thesis has meaning. It also seems that physicalism cannot be dismissed offhandedly as false (as according to the first horn of the dilemma) because formulated this way, the physicalist's hypothesis provides stipulations for its confirmation as well as its refutation.

The refutation condition of physicalism can be formulated as follows:

If the laws of complete physics are *divided* between two sets of objects that exhaust the macroscopic objects that exist in the actual world—then physicalism is *refuted*.¹⁷

¹⁷ The two sets must be divided with respect to the macroscopic non-physical properties of the objects in the actual world (as in the three examples discussed in Section 3) due to physics' pretense to completeness (discussed in Section 2). Otherwise, we could divide the objects in the world between charged objects and uncharged objects: charged objects require the physics of electromagnetism, while uncharged objects require only Newtonian mechanics. But we would not consider this a refutation of physicalism. Physicalism is the claim that fundamental physics, in principle, can explain all there is in the actual world—even what seems to us in the macroscopic world as not physical (such as the mental, the biological, the moral, and the aesthetic). The question about the unification of physics itself is a different question, separate from the question raised by Hempel's Dilemma (although interestingly related—see a brief discussion at the end of this section). For example, if physics discovers that there are two physically different categories of physical entities—such as

But in any other case that the laws of complete physics are not divided between any two sets of objects that exhaust the macroscopic objects that exist in the actual world—however strange and counterintuitive it may appear—we must admit that physics has succeeded to explain all there is in the world, and became complete *in a non-trivial way*.¹⁸

Let us consider the case that future-and-complete physics will assume special particles, say mental ones. According to the formulation I have proposed, this category of particles poses no dilemma *as long as the situation is consistent with the second premise*. That is, if these “extra” fundamental entities or forces are required and assist *not only* in explaining psychological phenomena (objects in set B), but also in explaining billiard balls (objects in set A), then this is consistent with the supervenience of everything there is in the world upon physics. In this case, the existence of mental entities does not necessarily challenge physicalism.

Moreover, we can imagine a case in which supervenience is violated. For example, in the future, *unique* explanations, laws, or entities of physical theory may be needed to explain mental phenomena (objects in set B). More precisely, the existence of explanations, laws, or entities will be necessary *in addition* to the laws and entities that are sufficient to explain objects in set A. In this case, the physicalist’s hypothesis would be refuted, and physicalism will be proven false.

The same is true for the other sets as well: we can imagine a case in which supervenience is violated, not only from a dualistic point of view. For example, in the future, *unique* explanations, laws, or entities of physical

matter and anti-matter—we would not consider this a refutation of physicalism. The refutation condition helps us argue that physicalism is not a trivial thesis, and that the subject matter in the refutation condition is macroscopic objects and not fundamental physical entities.

¹⁸ One can ask the following question: does this mean that, for all we now know, physicalism is false? After all, we seem to have a division between the physics of the micro and the physics of the macro. The answer is that this division is precisely one of the reasons (among others) demonstrating that contemporary physics is incomplete. Physics is uncomfortable with this division in its laws and tries to find a unified theory in order to resolve this division. And so, the answer to the question is negative: we know nothing about future and complete physics, and we cannot jump to conclusions regarding the falsity of physicalism based only on current physics.

theory may be needed to explain organic phenomena (objects in set B'). More precisely, the existence of explanations, laws, or entities will be necessary *in addition* to the laws and entities that are sufficient to explain objects in set A'. In this case, the physicalist's hypothesis would be refuted, and physicalism will be proven false. Supervenience can also be violated from an emergentist point of view. For example, in the future, *unique* explanations, laws, or entities of physical theory may be needed to explain emergent phenomena (objects in set B''). More precisely, the existence of explanations, laws, or entities will be necessary *in addition* to the laws and entities that are sufficient to explain objects in set A''. In this case, the physicalist's hypothesis would be refuted, and physicalism will be proven false.

My suggestion tackles the case in which a future unified physical theory is achieved in a trivial way. For instance, future and complete physics could simply be a conjunction of two theories: a physical theory and a psychological theory / a physical theory and a biological theory / a physical theory and a theory regarding complex systems. If a physicalist, a dualist, a vitalist, or an emergentist would examine the alleged unified physical theory carefully, they would discover that the laws of "the final and complete physical theory" are divided between two sets (depending upon which point of view the division of sets was made) of objects and phenomena in the world. For example, there are a few allegedly fundamental laws of physics that explain the organic or mental parts of the world, but not the non-organic or non-mental parts of the world. Likewise, there are a few allegedly fundamental laws of physics that explain the emergent aspects of reality but not the non-emergent aspects of reality. If this is the case, the physicalist, dualist, vitalist or emergentist would know that the alleged unified physical theory is unified in a trivial way and physics is not truly complete. Hence, the physicalist hypothesis is refuted.

Those who think this proposal for examining future unified physical theory is farfetched may consider that it is common in the literature to question the alleged successful unifications in theoretical physics achieved to date. In short, four basic physical forces exist in fundamental physics: electromagnetism, gravity, the strong nuclear force, and the weak nuclear force. Theoretical physics is attempting to produce a theory unifying these forces. Its aim is to demonstrate that there is only one fundamental force in the

universe. The first step in this unification programme has already been achieved: electromagnetism has been unified with the weak nuclear force in the electroweak theory. Next, the electroweak force is to be unified with the strong nuclear force by a grand unified theory (GUT). Finally, the GUT will be unified with gravity in a Theory of Everything [TOE; (Maudlin 1996)].

But philosophers and physicists dare to question this particular unification programme. For example, Rescher (1999) claimed that combining the four fundamental forces is insufficient because the ‘Theory of Everything’ must be holistic, and not simply an aggregation of forces. According to Maudlin (1996), the image of this unification programme has become so pervasive as to rank almost as a dogma. Why assume that these four forces are to be unified other than for purely aesthetic reasons?

Furthermore, Maudlin (1996) demonstrated that the unification of the weak nuclear force and the electromagnetic force was *forced* on those who were primarily engaged in seeking an adequate theory of the weak force. In fact, some theorists (for example Richard Feynman) deny that the electroweak theory displays any *real* unification of electromagnetism and the weak force: “it is not that at some point we had theories of the electromagnetic, weak, strong, and gravitational forces separately, and now we have managed to unify the first two. Rather, at some point, we *recognized the existence* of all four forces, and found that unification was needed to account for the weak force” (Maudlin 1996).

These examples show that even in the realm of fundamental physics we are asking questions regarding “real” and “holistic” unification, as opposed to just “aggregations” or “combinations”. My formulation of the physicalist’s hypothesis and its refutation condition is designed to answer similar considerations. One can see the resemblance between the point of view introduced in Hempel’s Dilemma and the skeptical point of view about unifications in physics itself.

5. Understand the dictation of the refutation condition

It seems that the advocate of Hempel’s Dilemma is familiar with physics’ pretense to completeness, and hence is troubled about the broadening of physics for the sake of its completion. Of course, most advocates of Hempel’s

Dilemma are concerned about the addition of mental aspects to physics in order to explain the mental parts of the world. Advocates from other perspectives can formulate similar dilemmas. For example, a hypothetical vitalist would state that physicalism is an ill-formed thesis: it is obvious that we do not yet have a complete physical explanation for every biological phenomenon (meaning that biology is still not fully reducible to physics), so it could be that for the sake of completeness, future physics will be expanded to include some non-reductive biological entities or forces. This makes physicalism, in the eye of this hypothetical vitalist, a false or unrefuted thesis.¹⁹

A more important point, perhaps, is that the physicalist can do nothing about it. The physicalist cannot set rules regarding the characteristics of the final entities and laws of physics. For this was the lesson physicalists learned from their predecessors: calling the thesis materialism was a mistake, resulting from the historical belief that physics deals only with matter. Today we know that the stuff of physics includes energies, forces, fields, and entities whose composition remains a mystery (particles, strings, membranes, or perhaps something else). A physicalist who understands the dependence of his thesis on the empirical content of physical science cannot characterize the nature of ultimate physical theory. To do so would be to ignore the end of materialism.

However, the physicalist can ensure that physicalism does not become a trivial thesis. One method for achieving this objective is to prohibit future splitting in the laws of physics between *any* two sets of objects that exhaust the macroscopic objects that exist in the actual world. For this reason, I constructed the physicalist's hypothesis as presented in Section 3. If a division would occur between the laws that govern objects and properties from set B, and the laws that govern objects and properties from set A, then we could assume that the additional laws are not part of physics, but were added from outside of physical theory in order to explain the

¹⁹ This is just a hypothetical example—I do not imply a necessary connection between reductionism and Hempel's Dilemma. One can argue for physicalism without having reductive explanations in hand, *and* it seems that if the argument of the dilemma is sound, then even having such reductions in hand would not allow one to claim that everything is physical. Hempel's Dilemma seems to be quite independent of any worries about the (lack of) success of reduction to physics.

mental phenomena. If a division would occur between the laws that govern objects and properties from set B', and the laws that govern objects and properties from set A', then we could assume that the additional laws are not part of physics, but were added from outside of physical theory in order to explain the biological. And if a division would occur between the laws that govern objects and properties from set B'', and the laws that govern objects and properties from set A'', then we could assume that the additional laws are not part of physics, but were added from outside of physical theory in order to explain the emergent phenomena of the world. In each of these cases, the physicalist's hypothesis would be refuted. Also, the dualist / vitalist / emergentist would be able to say that she was right all along because physics alone cannot explain the mental / organic / emergent aspect of the world.

For any other case, such as that of mental entities (or laws) in physics, which are needed in order to address or explain objects and properties from set A *and* set B, the physicalist should not care at all. The same is true in the case of a „vital force“ in physics needed to address or explain objects and properties from set A' *and* set B'. I argue that the physicalist should not care at all even if a new law is needed to address or explain objects and properties from set A'' *and* set B''. On the contrary, in the case of mental entities / vital force / emergence law, it could then be said that physics succeeded in uncovering the true structure of the world. For this is exactly what the physicalist assumes: it is the task of physics to discover the nature of the world and how that nature explains every manifestation in the world. This notion stems from understanding physicalism as a metaphysical hypothesis that is related tightly to the actual world. The physicalist cannot *veto* the character and nature of the content of physics. This is the crucial point of the dilemma.

The constraint regarding the un-splitting of laws that govern the objects and properties in the world is not external to the thesis. More importantly: this constraint is *general* and does not focus only on the mental aspect. These two aspects of the un-splitting constraint are opposed to the constraint “without mental entities in future and complete physics.” My constraint is generated from *within* the thesis. The physicalist should not be concerned with physics' articulation of the structure of the world. She

should care whether physics succeeds in uncovering that structure without being artificially completed by laws and entities.

6. What should we learn from Hempel's Dilemma?

I suppose most philosophers will still have difficulty accepting that physicalism can be true even if the future-and-complete physical theory includes fundamental mental entities (whatever they may be), and even if these entities meet the demands I introduced earlier (i.e. that those entities would be required for explaining phenomena and objects in both sets, A and B). Even the vast majority of physicalists may have difficulty accepting this idea. I realize how controversial this may sound: it can be argued that the notion of physicalism that I provide does not correspond to the use of the word 'physicalism' in the philosophical community (or at least in the way it is discussed within the mind-body literature). For typically, the physicalist rejects fundamental mentality.²⁰ But the typical physicalist also rejects fundamental biology and more fundamental stuff that are not physical. This prompts the question: what does count as 'physical'? And obviously, we have returned to the dilemma.

In any case, these proffered suggestions may be the real solutions to the dilemma, stemming from the elucidated hypothesis of physicalism and its refutation condition. However controversial, as physicalists we must note that the possibility that future and complete physics will contain "mental" entities, or "emergent" laws, or "vital" forces (whatever the meaning of these expressions may be in the future) is negligible. Nevertheless, the point is that being physicalists, we cannot a-priori rule out these possibilities completely. As argued by Spurrett (2017): "The fact of revision and revolution in the history of science, and the undoubted provisionality and incompleteness of science as we have it, do indeed tell against simply letting current science determine what the physical (or material) is for philosophical purposes." In my opinion, the crucial point is the physicalist's commitment

²⁰ Prelević (2018) discussed the concern that postulating consciousness at the fundamental level is not in accordance with standard classifications in the history of philosophy (for further discussion see Section 7).

to physics, derived from the understanding of the motivation for being a physicalist and the point-of-view introduced in the dilemma.

We may want to think about the following question: is the purpose of the project to provide a notion of physicalism that makes sense of the debate and clarifies what philosophers have in mind when they are saying ‘physicalism,’ or is it more like a *normative project*, providing this notion of physicalism as the one that we should use? I see myself as a physicalist from naturalistic-empirical arguments. But the dilemma regarding the notion of the word ‘physical’ and its boundaries has always troubled me. I am obviously not offering here a solution to the question “what physics is meant?” but that is exactly because of the empirical nature of science and physics.

Either way, from a physicalist point of view, how can we set boundaries to the notion physical and still hold that “everything is physical”? After all, if everything is physical, then the line between physics and chemistry is only an arbitrary convention that can be moved according to our needs (or, the line was set by us this way because of our physical brains that make us see the world divided in such and such manner), as is the line between chemistry and biology, and so on. From this point of view, the project of capturing the notion ‘physical’ is paradoxical in its essence and is bound to fail. Trying to capture an empirical notion such as ‘physical’ contradicts the essence of the empirical sciences.

7. A note regarding the other solutions to Hempel’s Dilemma

Dowell (2006) and Vicente (2011) criticize the use of *via negativa*, maintaining that it is unnecessary to include the mental in the definition of the physical out of skepticism regarding a priori truths in general. I agree with this view. My use of the *via negativa* in the physicalist’s hypothesis is only partial. I divided the world into two sets of objects, set A containing objects that do not possess mental properties, and set B containing objects that do possess mental properties. While this initial step is, in and of itself, use of *via negativa*, one should continue and elucidate the physicalist’s hypothesis to answer the dilemma. In other words: one should use the refutation condition of physicalism. As noted above, it is obvious that the dilemma can be articulated from *other* points of view. If the vitalist would ask “what is

physical?” and the answer would be *via negativa*, “not biological”, this surely is not a sufficient solution. If the physical is not mental, and not biological, and not aesthetic, and not moral, and so on—then what is the physical? Apparently, we are left with the answer: “the physical is physical”... and our dilemma remains unresolved. The crucial point is this: my suggested construction of the physicalist’s hypothesis is only a pragmatic step in the purpose of formulating a *refutation condition* of physicalism.

Furthermore, setting aside what to classify as the ‘*via negativa*,’ it is worth noting that before that term became current, some had urged pragmatic flexibility about what to exclude from the physical. Spurrett and Papineau (1999) took an overtly pragmatic line, saying “[which] completeness thesis you ought to be interested in thus depends on the purpose to which you want to put the causal argument.” (A ‘completeness thesis’ refers to a causal closure principle.) In later work, Papineau (2001) also urged flexibility about how to apply the ‘ideal physics as long as it does not include X’ template, saying: „The same point applies if we want to apply the causal argument to chemical, or biological, or economic states. As long as we can be confident that all non-chemical effects are fully caused by non-chemical (non-biological/non-economic...) states, then we can conclude that all chemical (biological/economic...) states must be identical with something non-chemical (non-biological/non-economic...)”²¹

However, the line suggested in (Spurrett and Papineau 1999) and in Papineau’s (2001) more pragmatic remarks, is not how most philosophers understand the ‘*via negativa*.’ Hence, my offer, rather than being radically new, picks up a neglected line of thought and shows how it has traction in the contemporary literature on Hempel’s Dilemma.

On a related point, it is worth noting a slightly obscure form of vitalism associated with Walter Elsasser (1958; 1962), according to which the laws of ‘physics’ (the science of the inanimate) were special cases of the laws of the animate. However, Elsasser was not a pan-psychist, although he may in some sense have been a pan-biologist or pan-vitalist. His relevance here is that his view suggests a different set of solutions for the template developed in this paper about laws and objects. Elsasser thought that there was

²¹ I thank an anonymous referee for this line of thought.

a complete science of all objects and that it had ‘biotonic’ laws. But it was not ‘physics,’ and he rejected what he called ‘materialism.’²² The trivial physicalist betting on future/ideal physics would, it seems, attach the label ‘physical’ onto the ‘biotonic’ laws. The non-mentalist proponent of the ‘via negativa’ would, perhaps, do so as well. However, it is not clear that Papineau (2001) would want to follow suit, and I believe it is clear he would say that in an Elsasser world, physics is not complete. However, in my opinion, and according to the formulation and the refutation condition I have offered for physicalism—it is possible that an Elsasser world is a world compatible with the physicalist hypothesis.

As for both Ney’s (2008) solution and the suggestion to include some constraints in the future and complete physical theory, I have shown that there is a possibility that physicalism actually permits such “mental stuff” (or “vital stuff,” or any other stuff) in fundamental physics, as long as such inclusion is consistent with the physicalist’s hypothesis. Hence, it seems that there could be a hypothetical case in which both physicalist and dualist would be saying that they had been right all along. They would both be right, in a way. The physicalist would have been right when claiming that physics will uncover the true nature of the world and that all ‘God’ had to do was to create the entities and laws, and everything else would come into existence automatically. The dualist (or panpsychist) would also have been right since mentality is in the fundamental stuff of the world.²³

Dowell (2006) regarded the physical not as metaphysical but as methodologically empirical. Dowell suggested tying physicalism’s ontological commitments to our best methods for justifying our beliefs about the natural world: TOE should be a theory with the hallmarks of scientific theories.

²² For more about Elsasser’s view see (Bronowski 1970) and (Gatherer 2008).

²³ I admit that this suggestion raises concern: maybe the mental is fundamental, for example in the form of panpsychism. Then it would not be true that we need a separate set of laws to govern the mental and the physical and, in my view, physicalism would not be refuted. I suggest that this is a bullet that we might have to bite. It is not for us to decide, a priori, what physics will end up looking like, or what the world will turn out to be. It is certainly true that we cannot say, a priori, what physics will contain. And one does not want this to turn into a mere verbal dispute about the use of the term “physical.”

Dowell (2006) also opted for a future physical theory that postulates consciousness at a fundamental level in accordance with physicalism, without making this view trivial or empty. Prelević (2018) discussed the concern that postulating consciousness at the fundamental level is not in accordance with standard classifications in the history of philosophy and, in fact, may allow classifying philosophers like Descartes, Leibniz and perhaps Chalmers as physicalists, which seems implausible. However, I believe the refutation condition I offer may help clarify this concern. Although I agree with Dowell, I believe that, in a way, my view expands Dowell's view by providing a refutation condition of physicalism that does not require a strict notion of the physical that opposes the empirical essence of physics. Furthermore, using my refutation condition, not all future TOEs postulating consciousness at the fundamental level will be in accordance with physicalism—even if they have the hallmarks of scientific theories. In this respect, my view is distinguished from the view of Dowell (2006).

Of course, if at the end of physics we find ourselves considering the “Theory of Everything” from the dualist point of view, and use my refutation condition to ensure that the unification is not trivial, there will be an inherent risk. Emergentists about complex entities will then classify as physicalists, for if hurricanes and cells make it into set A, then no matter what view you have about them, your view will ipso facto count as physicalist. The same risk is inherent for some vitalists, for if viruses and bacteria become part of set A, then no matter what view you have about them, your view will ipso facto count as physicalism.

My answer to this quandary is this: it is possible that as soon as we have the TOE, we will need to start using my refutation condition from *several* points of view to ensure that the unification was not achieved in some trivial way. We will have to examine carefully the laws of the final physical theory by dividing the world into various sets (such as set A' and B', and A'' and B'')²⁴ to ensure that the laws of physics are not divided. We might find, for example, that physics has succeeded in explaining mentality but failed to explain other complex entities. In that imaginary scenario, although (perhaps) there will be no fundamental mentality in physics, we will still be

²⁴ There can be many other possible sets in future science of which we are unable to conceive at this point in time.

able to argue (using my refutation condition) that physicalism is false from an emergentist's point of view.

A good example of a new point of view is Integrated Information Theory (IIT), one of today's most influential theories regarding consciousness. Applying my refutation condition of physicalism to the case of IIT can be helpful in comparing the approach in my paper to other prominent solutions to the dilemma.²⁵ According to IIT, consciousness is integrated information, meaning that the quantity of consciousness corresponds to the amount of integrated information generated by a complex of elements, and the quality of experience is specified by the set of informational relationships generated within that complex (Tononi 2008). This theory has ontological consequences: according to IIT, integrated information exists as a fundamental quantity—as fundamental as mass, charge, or energy. Since consciousness is the same thing as integrated information, IIT argues that consciousness itself is a fundamental property (Tononi 2008). So, concerning Hempel's Dilemma, is IIT compatible with physicalism? Alternatively, in the event that we acquire empirical justification for IIT in the future, we will conclude that physicalism is a false thesis?

It seems that according to the *via negativa* we can answer immediately, even without waiting for empirical justification: since according to the *via negativa* we define the physical as non-mental, if we find out that indeed the physical can also be mental in some cases (i.e., integrated information), then physicalism is false. But when applying my refutation condition to this particular case, the answer is not so straightforward. For in the case of IIT, we can imagine a situation in which supervenience is violated. For example, in the future, *unique* explanations, laws, or entities of physical theory may be needed to explain computational properties or computational phenomena. More precisely, explanations, laws, or entities will be necessary *in addition* to the laws and entities that are sufficient to explain objects without computational properties. Of course, the opposite case is also possible: let us consider the case that future-and-complete physics will assume laws regarding these so-called fundamental integrated information properties

²⁵ It would also be interesting to apply my refutation condition of physicalism to some dualistic interpretation of quantum mechanics [for example see (Barrett 2006)]. However, this is beyond the scope of the current paper.

(whatever they will be). According to the formulation I have proposed, this category of fundamental laws and properties poses no dilemma, if these „extra“ fundamental entities / forces / laws / properties are required and assist *not just* in explaining brains and computers, but also in explaining tables, chairs, and rocks, then this is consistent with the supervenience of everything there is in the world upon physics. In this case, the existence of integrated information properties in the fundamental level of the world does not necessarily challenge physicalism.

Developing this example is beyond the scope of the current paper. Nevertheless, this example of IIT emphasized the point that the other solutions to the dilemma (such as the *via negativa* and futurism) are quick in judging a-priori what we should find in the fundamental level and what we should not while ignoring the empirical characteristics of physics itself. I believe that my solution leaves both options open while not avoiding the problem or ignoring the possibility that future physics will be entirely different than we can imagine.

8. Conclusion

This paper joins the ongoing argument over how best to respond to what has become known as ‘Hempel’s Dilemma,’ the choice facing physicalists over whether, when they say “everything is physical” they mean current physics or ideal/future physics. I defend a version of the respondents saying that the ‘future/ideal’ physics option can be selected without the consequence of making physicalism trivial. The triviality consequence is avoided because of the inclusion of a ‘refutation condition.’ This refutation condition is a version of the ‘*via negativa*’ (standardly taken to be a stipulation that the physical not include the mental), albeit one that is specified and worked out in a distinctive way. Thus, the solution offered here can be described as a new member of an established family of strategies: picking up a neglected line of thought and showing how it has traction in the contemporary literature on Hempel’s Dilemma.

Hempel’s Dilemma arises due to the most common catchphrase of physicalism “everything is physical,” and our difficulty stabilizing the extension of the concept physical. It seems that the advocate of Hempel’s Dilemma is

concerned that physics will broaden itself artificially to become complete. Thus, the dilemma is not about distrust of physicalism. The dilemma presents a concern about not trusting *physics*. Moreover, the physicalist is powerless to prevent physics from artificially broadening for the sake of becoming complete.

I have suggested that in order to confront the dilemma, we must first clarify the physicalist's hypothesis, thereby dismissing ambiguous definitions. What the physicalist can do is to ensure that physicalism is not a trivial thesis by strengthening the physicalist's hypothesis to address the concern raised by Hempel's Dilemma. The hypothesis suggested is that in future and complete physics there will be no splitting in the laws between the set of objects possessing mental properties and the set of objects not possessing mental properties (or any other two sets according to which we choose to divide the objects in the world). If a splitting took place, both dualist (or any other opponent to physicalism) and physicalist would know that physics has failed in its attempt to become complete and that physicalism is refuted.

My approach differs from others not only because of its suggested refutation condition for physicalism but also through its direct attention to vitalism and emergence. The recent tendency to focus on Hempel's Dilemma as an issue for physicalism about the mind neglects the importance of vitalism and emergence for the history of physicalism and for thinking about how to distinguish the physical from the non-physical.

By providing a more accurate answer to Hempel's Dilemma—a general refutation condition of physicalism—I bring to light the understanding about that to which the physicalist is committed. Even if this commitment is but a fraction more than what appears in the dilemma, it would be sufficient to show that physicalism is not a trivial thesis. In other words: there IS a question of physicalism.

Acknowledgements

I am very much indebted to Meir Hemmo for many stimulating discussions of these matters. I am also grateful to Yael Raizman-Kedar, who provided extensive written feedback on an earlier draft, and to David Buzaglo, Aya Evron, and the attendees of the 2016 "Cognition Research Seminar" at Haifa University for a helpful discussion. Many thanks to the anonymous reviewers for *Organon F* for valuable comments and suggestions provided.

References

- Barrett, Jeffrey A. 2006. "A Quantum-Mechanical Argument for Mind–Body Dualism." *Erkenntnis* 65 (1): 97–115. <https://doi.org/10.1007/s10670-006-9016-z>
- Bishop, Robert C. 2006. "The Hidden Premiss in the Causal Argument for Physicalism." *Analysis* 66 (289): 44–45. <https://doi.org/10.1111/j.1467-8284.2006.00588.x>
- Bokulich, Peter. 2011. "Hempel's Dilemma and Domains of Physics." *Analysis* 71 (4): 646–51. <https://doi.org/10.1093/analys/amr087>
- Braddon-Mitchell, David, and Frank Jackson. 1996. *The Philosophy of Mind and Cognition*. Cambridge, MA: Blackwell Publishers.
- Bronowski, Jacob. 1970. "New Concepts in the Evolution of Complexity." *Synthese* 21 (2): 228–46. <https://doi.org/10.1007/BF00413548>
- Chalmers, David J. 2006. "Strong and Weak Emergence." In *The Re-Emergence of Emergence*, edited by Philip Clayton and Paul Davies, 244–56. Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199544318.003.0011>
- Crane, Tim, and D. H. Mellor. 1990. "There Is No Question of Physicalism." *Mind* 99 (394): 185–206. <https://doi.org/10.1093/mind/XCIX.394.185>
- Crook, Seth, and Cart Gillett. 2001. "Why Physics Alone Cannot Define the 'Physical': Materialism, Metaphysics, and the Formulation of Physicalism." *Canadian Journal of Philosophy* 31 (3): 333–60. <https://doi.org/10.2307/40232121>
- Dowell, Janice L. 2006. "The Physical: Empirical, not Metaphysical." *Philosophical Studies* 131 (1): 25–60. <https://doi.org/10.1007/s11098-005-5983-1>
- Elsasser, Walter M. 1958. *The Physical Foundation of Biology: An Analytical Study*. London, NY, Paris and Los Angeles: Pergamon Press.
- Elsasser, Walter M. 1962. "Physical Aspects of Non-Mechanistic Biological Theory." *Journal of Theoretical Biology* 3 (2): 164–91. [https://doi.org/10.1016/S0022-5193\(62\)80013-7](https://doi.org/10.1016/S0022-5193(62)80013-7)
- Fiorese, Raphaël. 2015. "Stoljar's Dilemma and Three Conceptions of the Physical: A Defence of the Via Negativa." *Erkenntnis* 81 (2): 201–29. <https://doi.org/10.1007/s10670-015-9735-0>
- Gatherer, Derek. 2008. "Finite Universe of Discourse. The Systems Biology of Walter Elsasser (1904–1991)." *Open Biology Journal* 1: 9–20. <https://doi.org/10.2174/1874196700801010009>
- Gillett, Carl, and D. Gene Witmer. 2001. "A 'Physical Need': Physicalism and the Via Negativa." *Analysis* 61 (4), 302–09. <https://doi.org/10.1093/analys/61.4.302>
- Hemmo Meir, and Orly R. Shenker. 2012. *The Road to Maxwell's Demon*. Cambridge: Cambridge University Press.
- Hempel, Carl Gustav. 1966. *Philosophy of Natural Science*. Englewood Cliffs, NJ: Prentice-Hall.

- Kim, Jaegwon. 2006. "Emergence: Core Ideas and Issues." *Synthese* 151 (3): 547–59. <https://doi.org/10.1007/s11229-006-9025-0>
- Kripke, Saul. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1983/1999. "New Work for a Theory of Universal." Reprinted in *Papers in Metaphysics and Epistemology*, 8–55. Cambridge: Cambridge University Press.
- Maudlin, Tim. 1996. "On the Unification of Physics." *The Journal of Philosophy* 93 (3): 129–44.
- Meehl, Paul E., and Wilfred S. Sellars. 1956. "The Concept of Emergence." In *Minnesota Studies in the Philosophy of Science*, vol. 1, edited by Herbert Feigl and Michael Scriven, 239–52. Minneapolis: University of Minnesota Press.
- Montero, Barbara. 2001. "Post-Physicalism." *Journal of Consciousness Studies* 8 (2): 61–80.
- Montero, Barbara, and David Papineau. 2005. "A Defense of the Via Negativa Argument for Physicalism." *Analysis* 65 (287): 233–37. <https://doi.org/10.1093/analys/65.3.233>
- Ney, Alyssa. 2008. "Physicalism as an Attitude." *Philosophical Studies* 138 (1): 1–15. <https://doi.org/10.1007/s11098-006-0006-4>
- Papineau, David. 2001. "The Rise of Physicalism." In *Physicalism and Its Discontents*, edited by Carl Gillett and Barry Loewer, 3–36. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511570797.002>
- Papineau, David. 2016. "Naturalism." *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/win2016/entries/naturalism/>.
- Prelević, Duško. 2017. "Hempel's Dilemma and Research Programmes: Why Adding Stances Is Not a Boon." *Organon F* 24 (4): 487–510.
- Prelević, Duško. 2018. "Physicalism as a Research Programme." *Grazer Philosophische Studien* 95 (1): 15–33. <https://doi.org/10.1163/18756735-000023>
- Redhead, Michael. 1996. *From Physics to Metaphysics*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511622847>
- Rescher, Nicholas. 1999. *The Limits of Science*. Pittsburgh: University of Pittsburgh Press. <https://doi.org/10.2307/2181922>
- Robinson, William. 2015. "Epiphenomenalism." *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/fall2015/entries/epiphenomenalism/>.
- Spurrett, David. 2017. "Physicalism as an Empirical Hypothesis." *Synthese* 194 (9): 3347–60. <https://doi.org/10.1007/s11229-015-0986-8>

-
- Spurrett, David, and David Papineau. 1999. "A Note on the Completeness of 'Physics'." *Analysis* 59 (261): 25–29. <https://doi.org/10.1111/1467-8284.00144>
- Stoljar, Daniel. 2010. *Physicalism*. London and New York: Routledge.
- Stoljar, Daniel. 2017. "Physicalism." *The Stanford Encyclopedia of Philosophy* (Winter 2017 Edition), edited by Edward N. Zalta. URL = <https://plato.stanford.edu/archives/win2017/entries/physicalism/>.
- Teller, Paul. 1992. "A Contemporary Look at Emergence." In *Emergence or Reduction?*, edited by Ansgar Beckermann, Hans Flohr and Jaegwon Kim, 139–53. Berlin: De Gruyter.
- Tononi, Giulio. 2008. "Consciousness as Integrated Information: A Provisional Manifesto." *The Biological Bulletin* 215 (3): 216–42. <https://doi.org/10.2307/25470707>
- Vicente, Agustín. 2011. "Current Physics and 'the Physical'." *The British Journal for the Philosophy of Science* 62 (2): 393–416. <https://doi.org/10.1093/bjps/axq033>
- Wilson, Jessica. 2006. "On Characterizing the Physical." *Philosophical Studies* 131 (1): 61–99. <https://doi.org/10.1007/s11098-006-5984-8>
- Worley, Sara. 2006. "Physicalism and the Via Negativa." *Philosophical Studies* 131 (1): 101–26. <https://doi.org/10.1007/s11098-005-5985-z>

Do We Share a Language? Communitarism and Its Challenges

Matej Drobňák*

Received: 20 July 2018 / Accepted: 5 December 2018


Abstract: The idea that natural languages are shared by speakers within linguistic communities is often taken for granted. Several philosophers even take the notion of shared language as fundamental and that allows them to use it in further explanations. However, to justify the claim that speakers share a language, it should be possible to demarcate the shared language somehow. In this paper, I discuss: A) the explanatory role which the notion of shared language can play, and B) a strategy for demarcating shared languages from within the linguistic production of speakers. The aim of this paper is to show that the indeterminate nature of meaning in natural languages problematizes the intuitive idea of natural languages as shared.


Keywords: Communitarism; indeterminacy of meaning; metasemantics; pragmatics; shared language.

1. Introduction

We often take the notion of shared language for granted. We talk about English-speaking countries or German grammar and at the New Year we resolve to improve our Spanish or to learn yet another language. To see

* University of Hradec Králové

 Department of Philosophy and Social Sciences, Philosophical Faculty, University of Hradec Králové, Rokitsanského 62, 500 03 Hradec Králové, Czech Republic

 matej.drobnak@gmail.com



a natural language as something what we can (and do) share with other speakers is very intuitive for lay persons as well as for philosophers.

In this paper, I use the label ‘communitarism’ for those philosophical approaches to natural languages that take the notion of shared language as fundamental. For example, Lewis (1969), Brandom (1994, 2000), and Peregrin (2014a, 2014b) argue that a shared language is an outcome of some intra-group processes and Borg (2004) argues that the sharing of natural languages is an outcome of innate language modules.¹ In general, proponents of communitarism believe that natural languages are shared in a sense that members of a community share one set of meaningful expressions and syntactic rules and that speakers play an important role in maintaining shared languages by using them in communication.

However, as Davidson (1986, 1994) noted, if we look at how communication works, we can notice that the way we use sentences or expressions is not always in accordance with their pre-learned, and thus shared, meanings. In other words, the linguistic production of speakers exhibits variability. This variability is caused by the fact that what a speaker means by uttering a sentence depends partially on her general knowledge and beliefs as well as on the broader circumstances of the conversation in which the sentence is uttered.

Many communitarists admit that the notion of shared language alone does not suffice to explain what makes communication successful. Because of that Lewis (1969, 1979) and Brandom (1994, 2000) stick to the notion of scorekeeping in language games, and Borg (2004) sticks to the distinction between semantic and pragmatic features of natural language processing.

However, if communication includes processes beyond the simple application of a shared language, the question arises what criteria of individuation should be used for the demarcation of shared languages from within the variable linguistic production of speakers. If communitarists believe that speakers within a community share a language and maintain it over time, then it should be possible to track and demarcate the shared language (the shared set of meaningful expressions and syntactic rules) from their linguistic production.

¹ Other philosophers, such as Dummett (1986, 1994) or Weiss (2010), use the notion of shared language as an important part of their argumentation.

In this paper, I discuss one of the most influential strategies which communitarists can adopt to demarcate shared languages—the strategy for coherently maintaining the intuitive idea of natural languages as shared and the idea that the linguistic production of speakers exhibits variability. In short, a communitarist can claim that *not all the aspects of the linguistic production of speakers* are relevant for demarcating what the shared language of a community is.² One way to distinguish between relevant and irrelevant aspects of linguistic production is to stick to some version of semantic-pragmatic distinction. In this paper, I discuss the currently predominant version of this distinction—minimal semantics as advocated by Borg (2004, 2012).

The aim of this paper is to show that the strategy faces serious problems. The problems lie in the fact that the meanings of expressions in natural languages are indeterminate, as well as in its two-step model of communication and understanding, which is currently criticized on empirical grounds (Cosentino et al. 2017).

In the first part of the paper, I will discuss why a coherent view of natural languages should take into consideration the linguistic production of speakers, as well as how the variability of linguistic production challenges the intuitive idea of natural languages as shared. In the second part of the paper, I will present the strategy mentioned above in detail.

I believe that the intuitive idea of natural languages as shared is so pervasive that it is often accepted without explicit reflection. Nevertheless, the aim of the paper is not to argue that we should abandon the notion of shared language, but to point out weaknesses of the strategy for the demarcation of shared languages and to open up a discussion about future improvement. Any alternative to communitarism currently under discussion

² In some sense, we can say that the strategy is an allusion to Chomsky (1965) and his distinction between competence and performance, because only some performances (actual acts of uttering a sentence) are acceptable as relevant data for the demarcation of shared languages. Note that the competence-performance distinction has been heavily criticized (Labov 1971, 468; Noonan 1999, 21) for its arbitrary preference of some data over others. Thus, if communitarists rely on the restrictions on “performance,” then similar objections should apply to them as well.

brings more problems than solutions and their prospects for providing a coherent view of natural languages are, at least for now, poor.³

2. Common ground

I start from a simple assumption: any philosophical account of natural languages should take into consideration how speakers understand expressions and sentences and, subsequently, how they use them in communication. The linguistic production of speakers, as an outcome of their competence, should be of *some* relevance for any philosophical explanation of natural languages simply because natural languages are those languages which developed naturally within communities of speakers and are used by those speakers in communication.

On the other hand, it also sounds intuitively acceptable that linguists, semanticists, or philosophers of language should abstract from the actual linguistic practices of speakers when providing explanations of how natural languages work. There are many reasons for abstracting—including the defectiveness and sloppiness of the actual linguistic production of ordinary speakers. Despite that, abstractions cannot be completely arbitrary. There must be some connection between the results of abstractions and the linguistic production of speakers. Otherwise there would be no justification that those results fit a particular natural language.⁴ In other words, there

³ Semantic holism as a version of an individualistic approach based on the notion of idiolects is an alternative (Rapaport 2000, 2003; Pollock 2014). See (Drobňák 2018) for a discussion of why Quine (1960) is a proponent of an approach which relies on the notion of idiolects as well. Another alternative could be Ludlow (2014) and his idea of microlanguages.

⁴ Such a result of an abstraction can be, for example, a semantic model of a language. A semantic model represents a natural language by means of a formal language. Formal languages abstract from natural languages by interpreting (hidden) structures of sentences of a natural language as precise and well-defined structures of a formal language. I believe that formalization of natural languages can help us to recognize some reasons for the variability of linguistic production, e.g. to recognize specific syntactic features of indexical expressions. But it does not take into account many other reasons, e.g. the role of the intentions of speakers in the variability of

must be some criteria of individuation for shared languages which tell us how to abstract from the linguistic production of speakers.

Because of that, the idea that linguists, semanticists, or philosophers of language abstract from *some aspects* of the linguistic production of speakers sounds more reasonable. The question of which aspects are relevant and which should be overlooked is then decided by criteria of individuation. In such a case, the actual linguistic production of speakers is considered to be a reliable source of data about natural languages, but we are finical in delimiting which aspects of the linguistic production count as a reliable source.

To sum up, any coherent view of natural languages should somehow reflect upon the actual linguistic practice; otherwise it is not clear what makes it about a natural language. In particular, since the actual linguistic production is performed by speakers in communication, any coherent view of natural languages should be able to give compatible answers to three questions:

- a) What natural languages are?
- b) What role particular speakers have in maintaining meaningful expressions in natural languages?
- c) How communication relates to the previous questions?

2.1. The challenge

The biggest challenge in providing the answers for a), b), and c) is that the traditional and very intuitive idea of language does not fit very well with the way in which communication works.

The traditional view in semantics or philosophy of language is that language is a set of meaningful expressions and syntactic rules. Such a view of language is implied if the principle of compositionality is accepted. Standard approaches that aim at delivering semantic models of languages assume that the meanings of words and syntactic rules are sufficient for composing the meanings of sentences (usually understood as truth conditions). The lexicon

their linguistic production. The idea will be further discussed in Section 3 under the label of syntactically-triggered context sensitivity.

of a language, i.e. the set of meaningful expressions, serves as a stock of building blocks for compositionality. If I learn the meaning of an expression, I can use it (together with other expressions and syntactic rules) to compose what is basically an infinite number of sentences. The role of expressions as building blocks for compositionality is facilitated by the fact that the majority of expressions have determinate and context-invariant meanings.

Such a view of languages has a very high explanatory potential. First of all, a language that mostly consists of expressions with one determinate and context-invariant meaning is easy to learn and share. If we assume that different speakers can acquire and share the meanings of expressions (or concepts as their mental representations), we can easily answer all three of the above questions. We can explain what a natural language is by saying that it is the set of meaningful expressions and syntactic rules which is shared by some speakers. By learning the shared language, speakers become competent and maintain the language for subsequent generations. Then, a group of people counts as a linguistic community if and only if almost all its members share the same language. In the same way, we can say that different speakers understand each other because they share the pre-learned language and their communication is successful because they ascribe the same meanings to the same expressions.

The biggest problem of this approach is that such an explanation of communication does not have much support if we look at how it actually works. Davidson (1986, 1994) argues that the way in which understanding is reached shows it to be untenable that all speakers simply assign the same pre-learned meanings to the same expressions in communication. Such a view of understanding is untenable, because how we understand a sentence may be influenced by contextual cues present in a conversation. Cues can be intentionally incorporated into conversation by a speaker, or they might be a result of the accidental circumstances in which a conversation takes place. As Davidson (1986, 439) demonstrates through Donnellan's (1968) use of the sentence 'There's glory for you,' even this sentence can be understood as 'There's a nice knockdown argument for you' if the conditions are right. Because of these conversational shifts in meaning, the linguistic production of speakers exhibits variability.

However, the point I want to emphasize is not about the explanation of communication. As I mentioned above, communitarists often admit that the explanation of successful communication requires more than just a simple application of a shared language. Rather, the point I want to emphasize is about demarcating the boundaries of shared languages. If our linguistic production (the way we use expressions) is a source of data for demarcating shared languages and, at the same time, there is a realm of linguistic production that exhibits variability, then the variability of the data indicates that expressions do not have determinate and context-invariant meanings and so it problematizes the possibility of demarcating shared languages.

The variability of linguistic production leads Davidson to the conclusion that there is nothing that corresponds to the standard view of language as a set of meaningful expressions and syntactic rules and we should abandon it. He claims that “there is no such thing as a language, not if a language is anything like what many philosophers and linguists have supposed” (Davidson 1986, 446) and he believes that “we must give up the idea of a clearly defined shared structure which language-users acquire and then apply to cases” (Davidson 1986, 446). I agree that the argument is valid, but only under the assumption that all linguistic production is taken to be a relevant source of data about shared languages. If the circumstances of particular conversations influence our linguistic production and all linguistic production is taken to be a relevant source of data about natural languages, then communitarists are losing the demarcation criterion for what counts as a shared language.

Rejecting the assumption that all linguistic production is a reliable source of data about natural languages might help communitarists to avoid Davidson’s conclusion, but the variability of linguistic production (as a fact about natural languages and their use) still poses a challenge for them. The challenge for communitarists is to give clear criteria for which aspects of the linguistic production of speakers count as a reliable source of data for demarcating shared languages and which aspects should be considered to be irrelevant. If communitarists want to preserve the notion of shared language, then they have to explain the existence of the variability of linguistic production along with the existence of shared languages.

3. Communitarism

One strategy for avoiding the challenge is to bite the bullet and to accept that the variability of linguistic production shows that the meanings of expressions in natural languages are, in some sense, variable. Biting the bullet does not necessarily mean the loss of the notion of shared language. The idea might be that, even though there are several ways in which an expression is used in communication, the ways are well recognized and shared by speakers within a community. The context-invariant meaning can be explicated as a compound of several contextual values and the shared language as a set of expressions with complex and variable meanings and syntactic rules. Such a view of meaning is sometimes labelled a ‘rich meaning approach’ but, as far as I know, this approach to meaning and shared languages has not been spelled out in detail so far.

The biggest problem of this strategy is that it is not clear whether meaning understood in this way can be compositional and thus whether creating a semantic model of a language would be possible. Another problem of this approach is that if meanings are complex compounds, acquiring such languages would be much more demanding (and probably almost impossible). Even if this strategy allows communitarists to save the notion of shared language (in the new sense), such a notion of shared language would not be able to play the same explanatory role as the standard notion of shared language was supposed to play—causing new complications and problems that must be solved.

3.1 Which aspects of linguistic production?

Another strategy for preserving the intuitive notion of shared language relies on setting a clear boundary between those aspects of linguistic production which are shared by all speakers and those aspects which can vary from speaker to speaker, from conversation to conversation. As stated earlier, the notion of shared language is often taken for granted without explicit reflection, so it is hard to find an explicit proponent of this strategy. However, I believe that the strategy can be naturally linked to Grice’s (1957, 1961) distinction between the semantic and pragmatic features of content. I believe that, if asked, many philosophers would stick to an explanation in

line with the Gricean distinction between semantics and pragmatics—we share a language with regard to semantics and the variability of our linguistic production is caused by pragmatics.

In particular, a proponent of communitarism can claim that only those aspects of our linguistic production which are relevant for semantic features of content serve as a reliable source of data about shared languages. If communitarists can succeed in demarcating which aspects of our linguistic production correspond to the semantic features of content, then they basically succeed in responding to the challenge. Even though this might not be its primary purpose, minimal semantics as advocated by Borg (2004, 2012) can serve as a very good background for accomplishing this task.

According to Borg (2004, 2012), formal semantics should deal with the literal meaning of sentence-types and expressions. More specifically, formal semantics should provide a model of a language that is able to state what each sentence of a language means, solely on the basis of the syntactic features of sentences and the semantic properties of its constituents (particular expressions of a language). Stating this standardly amounts to stating the truth conditions of sentences.⁵

What matters for semantic operations on a formal account just are the (local) syntactic properties of representations. So, on this kind of picture, grasp of meaning would seem to be in principle amenable to a (Turing-style) computational explanation. If, say, we treat grasp of literal linguistic meaning as the canonical derivation of truth conditions for sentences, for example, along the lines of Larson and Segal 1995, then semantic understanding can

⁵ According to Borg, this also allows the incorporation of syntactically triggered context-sensitive expressions. Overt context-sensitive expressions (e.g. demonstratives, indexicals, tensed expressions) may count as such. Syntactically triggered context-sensitivity amounts to cases in which context-sensitivity is somehow “built into” an expression. In other words, context-sensitivity is, in such cases, a matter of the syntactic properties of an expression—it is recognized by a hearer automatically just by hearing an expression, regardless of the broader context of a conversation. This sort of context-sensitivity is in striking contrast to different sorts of context-sensitivity such as conversational implicatures, which require knowledge of the context of a conversation in order to be recognized by hearers.

form part of a genuine language module, for this is clearly a function which is encapsulated and computational. Knowledge of meaning, on this kind of account, consists of knowledge of a proprietary body of information (the lexicon for the language) and knowledge of a set of rules operating only on that information, rules which consist of formal transformations of the data. (Borg 2004, 81)

Knowing the “proprietary body of information” requires knowing how to categorize objects under particular expressions of a language, i.e. knowing which expressions are related to which concepts. Note that, according to Borg, T-sentences map natural language sentences to “Mentalese,” so it makes sense to say that concepts are mental representations of meanings and so the categorization of objects under expressions is relevant for the semantic processing of sentences. If this is so, then referential aspects of our linguistic production can provide relevant data for demarcating literal meanings, as referential aspects of our linguistic production indicate how a speaker categorizes objects.⁶

Aspects of our linguistic production, which are not syntactically encoded, are a matter of what can be implied by uttering a sentence and belong to pragmatics. They are irrelevant for the semantic meaningfulness of expressions and syntactic processing of sentences. But most importantly, if minimal semantics is adopted as a background theory for communitarism, then we can say that the aspects of our linguistic production that are not syntactically encoded are irrelevant for the demarcation of shared languages.

More generally, there are two aspects of minimal semantics that make this theory appealing for communitarists:

⁶ “It would also fall within the purview of the language faculty to calculate the mental representation of the truth-condition for the natural language sentence ‘The cat is on the mat,’ where what is constructed is a language of thought sentence which exhibits connections to the external world just to the extent that the language of thought expressions out of which it is constructed exhibit such relations (to put it crudely, *since CAT hooks up to cats, and MAT hooks up to mats, the truth conditions for the natural language sentence ‘the cat is on the mat’ turns on how things stand with some cat and some mat.*”) (Borg 2004, 24, emphasis added)

- a) since the meaning of a sentence is syntactically encoded, it is possible to determine the literal meaning of a sentence (and so to understand its literal meaning) without any information about the circumstances of a conversation. This can be done solely on the basis of information about the literal meaning of lexical units and the syntactic structure of a sentence and this information is accessible by all speakers under all circumstances simply by hearing a sentence;⁷
- b) since minimal semantics is closely linked to the modular theory of mind,⁸ models provided by formal semantics are supposed to be models of a specific linguistic module which is responsible for semantic processing. In general, this module is considered to be innate and this gives us a reason to assume that different speakers process the literal meaning of sentences in the same way. In other words, different speakers ascribe the same meanings to the same expressions/sentences.

If a communitarist adopts minimal semantics, she can claim that the aspects of our linguistic production that are related to syntactically encoded truth conditions of sentences serve to demarcate what shared languages are. Since congruence on concepts (categorization of objects) matters for stating the truth conditions of sentences, only the referential aspects of our linguistic production are relevant for the meaningfulness of particular expressions. A language is then a set of meaningful expressions and syntactic rules with regard to the syntactically encoded truth conditions of sentences and referential aspects of the linguistic production of speakers. If we add the assumption that such a language is an outcome of our innate semantic processing module, we can expect all the speakers within a community to share a language. This allows communitarists to save the notion of shared language and use it in further explanations. For example, it can be used to state a demarcation criterion for linguistic communities: what makes a group of

⁷ “What minimalism specifies is the content a competent language user is guaranteed to be able to recover, given adequate lexical resources” (Borg 2012, 63).

⁸ Borg (2004) overtly discusses the modular theory advocated by Fodor (1983, 1998, 2000). Another modular approach can be found in (Chomsky 1971, 1975, 1986, 2000).

speakers a linguistic community is the fact that they all share a language in the aforementioned sense—that they all share semantic processing with regard to the truth conditions of sentences and they agree on the categorization of objects falling under particular expressions.

A modular theory of mind also answers what role particular speakers play in establishing and maintaining a natural language. As syntactic processing is innate, it does not require any special effort. We are all disposed to process sentences syntactically in the same way simply by virtue of being normal human beings. All we need to do is to show our successors which expressions refer to which objects in the world to the extent that they are able to grasp the corresponding concepts.⁹

As long as we agree on which objects fall under ‘blood,’ ‘hands,’ ‘the room,’ etc. in the sentence ‘The man over there left the room with blood on his hands,’ we can all (semantically) process and understand the sentence in the same way. Without doubt, much more can be implied by uttering the sentence (e.g. that the man is a killer), but minimal semantics allows communitarists to discriminate minimal standards that must be shared by all speakers and it allows communitarists to demarcate natural languages in terms of these minimal shared standards.¹⁰

3.2. Communitarism and communication

If communitarists adopt minimal semantics, the most natural view of communication may be a two-step model: semantic processing first, pragmatic

⁹ Allowing that the process of “grasping concepts” can be, at least partially, innately driven: “Finally, then, it seems that we might recognize a third way in which to understand what a module is, for we might view a module as a combination of our two previous accounts, so that a cognitive module comprises a proprietary body of information together with a proprietary set of rules or processes operating over that information. Again, both the rules and the representations they operate on are usually thought to be given innately; thus we have a model of a module as an innate and dedicated cognitive processor” (Borg 2004, 76).

¹⁰ This is not to say that Borg herself is a proponent of this view. My only assumption in this paper is that her view can be used to demarcate shared languages in such a way and that such a view might be intuitively appealing for many communitarists.

processing second.¹¹ When a hearer hears a sentence, she first processes it unconsciously via a semantic module. The result is that she understands what a sentence means (semantic understanding). In the next step, all the pragmatic information about the speaker and other circumstances intervenes and the hearer comes up with an interpretation of what the speaker might want to imply by uttering this sentence (pragmatic understanding). The reason why this approach to communication might be appealing for communitarists is that according to this view the notion of shared language is necessary for the explanation of how communication works. According to this proposal, pragmatic processing is only possible with background semantic processing. To reach a pragmatic understanding, which is usually what we care about in communication, a hearer must be “on the same page” as a speaker with regard to the literal meanings of sentences. This requires that they both share a language with regard to the truth conditions of sentences and categorization of objects under particular expressions. If this is not the case, then the initial data required for pragmatic processing might lead the hearer astray.

To sum up, minimal semantics a) is able to preserve the notion of shared language by delimiting truth-conditional and referential aspects of linguistic production as relevant for the demarcation of shared languages and b) relies on the notion of shared language in the explanation of how communication works by postulating a congruent semantic understanding as a precondition for pragmatic understanding.

¹¹ This is the view held by Borg as well. In general, Borg does not think that formal semantics should be able to explain how communication works, and she does not aim at giving such an explanation. But by setting minimal semantics into a modular theory of mind, she sets the idea of minimal semantics into a broader view of how semantic and pragmatic aspects of understanding relate to each other. And this relation indicates a two-step model: “On the one hand, then, semantic knowledge is important and special—without it we would be robbed of the ability to interpret the meanings of words and sentences and thus linguistic communication would be impossible. Yet, from another perspective, semantic knowledge is quite unimportant and peripheral—without all the other kinds of knowledge we have, semantic understanding would be pretty much worthless” (Borg 2004, 263).

4. Problems of the strategy

There are two problems for communitarists adopting this strategy. The first problem is related to the minimal standards that must be globally shared by all speakers. The second problem is related to the two-step model of communication.

4.1. Global sharing

The requirement of a shared language with regard to the sharing of the meanings of particular expressions seems to be too strong to expect. The problem is that the meanings of many expressions in natural languages are not fully determinate and context-invariant.

The point about context-invariance can be demonstrated through examples of free pragmatic enrichment. On the basis of what a hearer might know or find out during conversations, her understanding of the verb ‘stop’ in the sentence ‘The policeman stopped the car’ may vary, depending on whether the policeman was standing in the road, sitting in the car, or chasing the car.¹² As Recanati argues, the circumstances of a conversation in such cases influence not only pragmatic aspects of content (i.e. what is implicated by the sentence) but also the meaning of the sentence, because each way of stopping the car (by issuing a proper signal, by depressing the brake pedal, or by firing a warning shot) is related to different truth conditions.

In this paper, I will put the topic of context-invariance aside and I will focus on the indeterminate nature of meaning in natural languages in detail. The indeterminate nature of meaning in natural languages is often flouted because it is usually taken to be a problem of a small number of expressions only, i.e. vague expressions. Vague expressions share one characteristic feature—objects categorized under them can be ranked on a scale ranging from those which certainly belong in a category to those which certainly do not. Even though there are some uncertain cases, we all have a clear idea of

¹² The example is a modification of an example from (Recanati 2004, 2010). The same point could be demonstrated by the example of painted leaves as discussed in (Travis 1997).

a scale on which we move, i.e. we have well-established and shared criteria of categorization. For example, most people would agree that the percentage of the surface of a head without hair or the density of hair are among the relevant criteria for ‘bald.’ Different speakers may diverge on how they actually set thresholds, but it does not necessarily mean that their concepts diverge as well, i.e. that they use different criteria for the categorization of objects.

I believe that the indeterminate nature of meaning in natural languages is much more widespread. First of all, it may concern any expression in a natural language—including those that are not standardly understood as vague. Basically, for any expression in a natural language we can find circumstances in which the application of a criterion of categorization is unclear or undecided. The reason is that our linguistic practice is adjusted in accordance with some standard conditions in which we apply criteria of categorization. However, we all experience unusual conditions from time to time. In conditions in which it is not clear which criterion of categorization should be applied the decision is often in the hands of the participants in a conversation. If this is so, then it opens up the possibility that different speakers make different decisions and so they use different criteria of categorization, i.e. they assign different (though probably similar) concepts to one expression. Note that this is not vagueness as it is standardly understood. The problem I am discussing here is that we do not know whether some criteria of categorization are relevant, while in the case of vagueness we know what the relevant criteria are (we know the scale) but we do not know the exact thresholds.

Let us demonstrate this through the case of an expression that is not standardly considered to be vague—‘actor.’ Most people would agree that an actor is a person whose profession is acting in films or television and the number of appearances in films or whether acting is their main source of income are among the relevant criteria of categorization.¹³ In 2011, Orlean published an article in *The New Yorker* about Rin Tin Tin, a movie star from the twenties. He was a real star of those times—he starred in more than 20 films made by Warner Bros. and received the Abraham Lincoln

¹³ At least, this is a definition of the term provided by the Oxford Dictionary. See: <https://en.oxforddictionaries.com/definition/actor>.

humanitarian award and the medal for distinguished service, and the Mayor of New York gave him a key to the city. He was at the peak of his career in 1929 when he received the most votes for the best actor for the Academy Award. The only trouble was that Rin Tin Tin was a German Shepherd and the members of the Academy decided that a dog could not win the prize for the best actor.

Note that the question whether a dog can be an actor is not a case of vagueness. It is not a matter of how we decide to set the thresholds on a standard scale—Rin Tin Tin was the main character of many films and he was paid for his acting. He was even famous for real acting, as opposed to merely appearing on the stage (he was able to build the atmosphere of a scene by his facial expressions and so on). The question was whether being a human being is a relevant criterion for the categorization of objects under ‘actor’ and there is no vagueness in that; there is no blurred area of problematic cases. And yet, there was no definite answer to this question.

The case can be interpreted in two ways and both of them undermine the idea of minimal shared semantic standards and semantic processing. First, we can say that the meaning of ‘actor,’ or a corresponding concept, was indeterminate before the voting and it was only after realizing this indeterminacy that different people made it a little more precise.¹⁴ If this was the case, then it is hard to say what sharing indeterminate meanings/concepts amounts to. How can we say that two speakers shared the same meaning of ‘actor’ if it was not clear what the meaning was? How can we decide whether a concept possessed by one speaker is the same as a concept possessed by another speaker if it is indeterminate which criteria of categorization are constitutive for the concept? A natural response to this worry would be to say that those speakers possessed similar concepts or that their understanding of the expressions partially overlapped. However, as far as I know, there is no viable theory of concept/meaning similarity currently under discussion.¹⁵

¹⁴ Note that different people made different decisions so if this was the case, then the term became ambiguous. For the voters, ‘actor’ could include dogs; for members of the Academy it could only include human beings.

¹⁵ See (Fodor and Lepore 1999) for a critical evaluation of Churchland’s (1986, 1993) notion of meaning similarity.

Another interpretation of the Rin Tin Tin case is that even before the case it was determinate whether dogs fell under the concept of ‘actor’ but the question did not arise.¹⁶ In such a case, the unusual circumstances forced people to compare their understanding of ‘actor’ with respect to the categorization of dogs. The people who voted for Rin Tin Tin believed that a dog could count as an actor; the members of the Academy were a rather more conservative in their criteria of categorization. The understanding of ‘actor’ within a linguistic community had been challenged and it uncovered discrepancies between the concepts possessed by different speakers and thus meanings assigned to the same expression. If this was the case, then clearly the idea that all the speakers within a community share a semantic understanding on the level of particular expressions, i.e. ascribe the same meanings to the same expressions, does not have much support.

The idea that expressions of natural languages are indeterminate is not new. Waismann’s (1945) idea of open texture goes in the same direction and Gauker (2017), as a current proponent of the idea of open texture, overtly argues that it is problematic to simply assume that we all share the same fully determinate concepts (even though in most cases our criteria of categorization deliver overlapping results).¹⁷ Wilson (1982) proposes a thought experiment that aims to demonstrate that our criteria of categorization are often influenced by accidental features of situations in which decisions are made.¹⁸ Ludlow (2014) argues that our criteria of categorization are dynamic, i.e. they can change from conversation to conversation.

The lesson to be learned from the Rin Tin Tin case is that there is never a guarantee that there is a special realm of semantic processing which is

¹⁶ This is certainly an oversimplification. At least, it had never received so much attention.

¹⁷ However, see (Shapiro 2006) for a critical discussion. Shapiro argues that open-texture should count as a kind of vagueness. As far as I can see, the discussion does not have a winner so far.

¹⁸ The thought experiment is about an airplane that has fallen into the jungle. According to Wilson, the decision as to whether a jungle tribe will consider the plane to be a strange house or a strange bird depends on whether they see the plane before its fall or find it after the fall.

shared by all speakers. Even in a case such as ‘actor,’ in which we usually assume the congruence of our concepts without any doubts, we can find differences among the members of a linguistic community.¹⁹ Note that the indeterminacy revealed by unusual circumstances is not the exclusive domain of the term ‘actor.’ The term ‘actor’ is not usually considered to be vague or non-standard in any other way. This suggests that this kind of indeterminacy might be a general feature of natural languages. If this is so, then it should not be a problem to find more such cases. To illustrate that this phenomenon is more common than we might think, I will present more examples.

Ludlow (2014) discusses the very similar case of whether Secretariat (a horse) can be an athlete. The discussion over Secretariat followed a very similar pattern to the one over Rin Tin Tin. Sport Illustrated placed Secretariat on the list of the best athletes of the last century. This decision sparked a public debate. Some people defended the choice and some people disagreed. From the linguistic point of view, this discussion can be understood as revealing discrepancies in the criteria for the categorization of objects under the term ‘athlete’ between different competent English speakers. Similarly, Johnson (2018) discusses in *The Conversation* how current advances in technology and science problematize our understanding of the term ‘meat.’ The question that Johnson poses in her article is whether lab-grown meat should also be considered to be meat. From the linguistic point of view, this is a question of whether the standard criteria of categorization for the term ‘meat’ apply to lab-grown meat as well. As in the last two cases, there are people who believe that lab-grown meat counts as meat and there are people who disagree. The changed circumstances (caused by advances in technology) reveal the indeterminacy of a term that was not considered to be indeterminate. It shows how different competent speakers may apply different criteria of categorization without knowing about it until challenged.²⁰

¹⁹ A similar point leading to a conclusion that decisions about the categorization of particular objects depend on accidental features of particular situations, and so it is hard to expect congruence among all speakers, was raised in (Wilson 1982, 2006).

²⁰ Similarly, we could interpret the discussion over the status of Pluto as a linguistic discussion over the criteria of categorization of the term ‘planet.’ As in the case

The most important point, however, is that we can never rule out the possibility that we will stumble upon such indeterminacy or differences in concepts/meanings for any expression in a natural language because we can never assess all the possible circumstances in which an expression can be used. For any expression, there is a possibility that there will be some circumstances which may reveal differences in categorization which have not been noticed before. If this is so, then the assumption that there is a special realm of semantics on the level of particular expressions, which is shared by all speakers, seems at least problematic and deserves more attention by any communitarist adopting this strategy.

4.2. Two-step model of communication

The second problem of this strategy is the two-step model of communication. The idea that we first semantically process what we hear and it is only after that that we start pragmatic processing has been undermined by recent empirical research. Werning and Cosentino (2017) and Cosentino et al. (2017) show that free pragmatic enrichment intervenes even during the early stages of the semantic processing of sentences.²¹ More specifically, free pragmatic enrichment helps us to modulate word meanings before we process a sentence semantically. The research was focused on the neurological activity of subjects during the processing of congruent vs. incongruent noun-verb combinations.²² More specifically, research teams tested how neurological activity depends on the context of a sentence. A context was presented to subjects as a short story and it served as information necessary for free pragmatic enrichment. Beside other combinations, the researchers also tested how neurological activity changes

of ‘meat,’ new circumstances have been caused by advances in science. Specifically, by the discovery of other planet-size objects in Kuiper belt.

²¹ Free pragmatic enrichment is a process in which the semantic understanding of a term is influenced by a context of a conversation before the semantic processing of a sentence is finished. The term ‘free pragmatic enrichment’ was coined by Recanati (2004, 2010). See the beginning of subsection 4.1 for an example and a short explanation of the term.

²² Cosentino uses funnel-pour as an example of congruent combination and funnel-hang (a coat) as an example of incongruent combination.

when we combine a congruent context²³ with congruent and incongruent noun-verb combinations and when we combine an incongruent context²⁴ with congruent and incongruent noun-verb combinations.

According to the two-step strategy, subjects should first semantically process the literal meaning of the whole sentence and only then should pragmatic processing take place. If this is so, then we can predict that the context of a sentence should not influence neurological activity related to the semantic processing of particular words (or noun-verb combinations). More specifically, the neurological activity at the time of 400 ms after hearing a verb (N400 component)²⁵ should be the same, regardless of the context. However, the research shows that the neurological activity after hearing a verb is significantly affected in those cases in which a congruent context is followed by an incongruent noun-verb combination and vice versa (incongruent context and congruent noun-verb combination).²⁶ In other words, the same noun-verb combination elicits different neurological activity in different contexts. If we assume that the neurological activity (N400 component) corresponds to the contribution of a particular word to the processing of the meaning of a sentence, then the difference shows that the context influences a word's semantic contribution before the semantic processing of a sentence is finished.

²³ A context inducing a congruent noun-verb combination. In the case of funnel-pour, that would be a context of standard procedures in a chemical laboratory or in wine cellar.

²⁴ A context inducing an incongruent noun-verb combination. In the case of funnel-hang (a coat), that would be a context of creative work at an art class.

²⁵ The phenomenon of neurological activity peaking at the time of 400 ms after semantically oriented stimuli was reported for the first time by Kutas and Hillyard (1980) and confirmed several times after that in different experimental settings (Baggio et al. 2008; Kutas and Hillyard 1984; Kutas et al. 1984; Kutas and Federmeier 2011)

²⁶ Neurological activity was measured by EEG as electrical activity in a specific region of the brain 400ms after hearing a word. The influence of context was associated with the difference in electrical activity of the same part of the brain triggered by hearing the same word in different contexts.

More research has to be done to find out how exactly we should interpret the N400 component,²⁷ but all the interpretations currently under discussion hold that it represents the contribution of particular expressions to the semantic processing of a sentence. Differences in neurological activity varying in accordance with different contexts indicate that pragmatic processing takes place even before the semantic processing of a sentence is finished and so it undermines the two-step model of understanding and communication. But if the two-step model of communication is undermined, then the idea of a shared language based on minimal shared standards is problematized as well. There seems to be no special realm of semantic processing on the level of particular expressions which would be shared by all speakers.

5. Conclusions

In this paper, I discussed one strategy which communitarists can adopt for coherently maintaining the idea of a natural language as shared and the idea that the linguistic production of speakers exhibits variability in communication. According to the strategy, only some aspects of the linguistic production of speakers are relevant for demarcating a shared language. This strategy can naturally be supported by some version of a semantic-pragmatic distinction if the semantic features of content are considered to be shared by all speakers. The first problem of this strategy is that the meanings of expressions in natural languages are indeterminate and so it is hard to say what sharing meanings of expressions, and thus sharing a language, might amount to. The second problem is its two-step model of communication and understanding, which is currently criticized on empirical grounds.

Acknowledgments

I would like to thank an anonymous reviewer for helpful suggestions and comments on an earlier version of this paper.

²⁷ See (Baggio and Hagoort 2011); (Brouwer and Hoeks 2013); (Hagoort et al. 2009) for discussion.

Funding

Work on this paper was supported by the joint Lead-Agency research grant between the Austrian Science Foundation (FWF) and the Czech Science Foundation (GAČR) “Inferentialism and Collective Intentionality,” GF17-33808L.

References

- Baggio, Giosuè, van Lambalgen, Michiel, and Peter Hagoort. 2008. “Computing and Recomputing Discourse Models: An ERP Study.” *Journal of Memory and Language* 59 (1): 36–53. <https://doi.org/10.1016/j.jml.2008.02.005>
- Baggio, Giosuè, and Peter Hagoort. 2011. “The Balance between Memory and Unification in Semantics: A Dynamic Account of the N400.” *Language and Cognitive Processes* 26 (9): 1338–67. <https://doi.org/10.1080/01690965.2010.542671>
- Borg, Emma. 2004. *Minimal Semantics*. Oxford: Oxford University Press. <https://doi.org/10.1093/0199270252.001.0001>
- Borg, Emma. 2012. *Pursuing Meaning*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199588374.001.0001>
- Brandom, Robert. 1994. *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Brandom, Robert. 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.
- Brandom, Robert. 2007. “Inferentialism and Some of Its Challenges.” *Philosophy and Phenomenological Research* 74 (3): 651–76. <https://doi.org/10.1111/j.1933-1592.2007.00044.x>
- Brouwer, Harm, and John C.J. Hoeks. 2013. “A Time and Place for Language Comprehension: Mapping the N400 and the P600 to a Minimal Cortical Network.” *Frontiers in Human Neuroscience* 7 (758): 1–12. <https://doi.org/10.3389/fnhum.2013.00758>
- Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Chomsky, Noam. 1971. “Recent Contributions to the Theory of Innate Ideas.” In *Philosophy of Language*, edited by John Searle, 121–29. Oxford: Oxford University Press.
- Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.
- Chomsky, Noam. 1986. *Knowledge of Language*. New York: Praeger.
- Chomsky, Noam. 2000. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511811937>

- Churchland, Paul M. 1986. "Some Reductive Strategies in Cognitive Neurobiology." *Mind* 95 (379): 279–309. <https://doi.org/10.1093/mind/XCV.379.279>
- Churchland, Paul M. 1993. "State-Space Semantics and Meaning Holism." *Philosophy and Phenomenological Research* 53 (3): 667–72. <https://doi.org/10.2307/2108090>
- Cosentino, Erica, Giosuè Baggio, Jarmo Kontinen, and Markus Wernig. 2017. "The Time-Course of Sentence Meaning Composition. N400 Effects of the Interaction between Context-Induced and Lexically Stored Affordances." *Frontiers in Human Neuroscience* 8 (813): 1–17. <https://doi.org/10.3389/fpsyg.2017.00813>
- Davidson, Donald. 1986. "A Nice Derangement of Epitaphs." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest Lepore, 433–46. Cambridge: Blackwell.
- Davidson, Donald. 1994. "The Social Aspect of Language." In *The Philosophy of Michael Dummett*, edited by Brian McGuinness and Gianluigi Oliveri, 1–16. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-8336-7_1
- Donnellan, Keith. 1968. "Putting Humpty Dumpty Together Again." *The Philosophical Review* 77 (2): 203–15. <https://doi.org/10.2307/2183321>
- Drobnák, Matej. 2018. "Quine on Shared Language and Linguistic Communities." *Philosophia* 46 (1): 83–99. <https://doi.org/10.1007/s11406-017-9916-y>
- Dummett, Michael. 1986. "A Nice Derangement of Epitaphs: Some Comments on Davidson and Hacking." In *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*, edited by Ernest Lepore, 459–76. Cambridge: Blackwell.
- Dummett, Michael. 1994. "Reply to Davidson: The Social Aspect of Language." In *The Philosophy of Michael Dummett*, edited by Brian McGuinness and Gianluigi Oliveri, 257–62. Dordrecht: Kluwer Academic Publishers. https://doi.org/10.1007/978-94-015-8336-7_13
- Fodor, Jerry. 1983. *Modularity of Mind*. Cambridge, MA: MIT Press.
- Fodor, Jerry. 1998. *Concepts: Where Cognitive Science Went Wrong*. Oxford: Oxford University Press. <https://doi.org/10.1093/0198236360.001.0001>
- Fodor, Jerry. 2000. *The Mind Doesn't Work That Way*. Cambridge, MA: MIT Press.
- Fodor, Jerry, and Ernest Lepore. 1999. "All at Sea in Semantic Space: Churchland on Meaning Similarity." *Journal of Philosophy* 96 (8): 381–403. <https://doi.org/10.5840/jphil199996818>
- Gauker, Christopher. 2017. "Open Texture and Schematicity as Arguments for Non-Referential Semantics." In: *Meaning, Context, and Methodology*, edited by Sarah-Jane Conrad and Klaus Petrus, 13–30. Berlin: DeGruyter.

- Grice, Herbert Paul. 1957. "Meaning." *Philosophical Review* 66 (3): 377–88.
<https://doi.org/10.2307/2182440>
- Grice, Herbert Paul. 1961. "The Causal Theory of Perception." *Proceedings of the Aristotelian Society, Supplementary Volume* 35: 121–52.
- Hagoort Peter, Giosuè Baggio, and Roel M. Willems. 2009. "Semantic Unification." In *The Cognitive Neurosciences*, edited by Michael S. Gazzaniga, 819–36. Boston, MA: MIT Press.
- Johnson, Hope. 2018. "Should Lab-Grown Meat be labelled as Meat When It's Available for Sale? *The Conversation*. Available at: <https://theconversation.com/should-lab-grown-meat-be-labelled-as-meat-when-its-available-for-sale-93129>
- Kutas, Marta, and Kara D. Federmeier. 2011. "Thirty Years and Counting: Finding Meaning in the N400 Component of the Event-Related Brain Potential (ERP)." *Annual Review of Psychology* 62 (1): 621–47.
<https://doi.org/10.1146/annurev.psych.093008.131123>
- Kutas, Marta, and Steven A. Hillyard. 1980. "Reading Senseless Sentences: Brain Potentials Reflect Semantic Incongruity." *Science* 207 (4427): 203–05.
<https://doi.org/10.1126/science.7350657>
- Kutas, Marta, and Steven A. Hillyard. 1984. "Brain Potentials during Reading Reflect Word Expectancy and Semantic Association." *Nature* 307 (5947): 161–63. <https://doi.org/10.1038/307161a0>
- Kutas, Marta, Timothy E. Lindamood, and Steven A. Hillyard. 1984. "Word Expectancy and Event-Related Brain Potentials during Sentence Processing." In *Preparatory States and Processes*, edited by Sylvan Kornblum and J. Renquin, 217–37. Englewood Cliffs, NJ: Erlbaum Press.
- Labov, William. 1971. "The Notion of 'System' in Creole Studies." In *Pidginization and Creolization of Languages*, edited by Dell Hymes, 447–72. Oxford: Cambridge University Press.
- Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.
- Lewis, David. 1979. "Scorekeeping in a Language Game." *Journal of Philosophical Logic* 8 (3): 339–59. <https://doi.org/10.1007/BF00258436>
- Ludlow, Peter. 2014. *Living Words: Meaning Underdetermination and the Dynamic Lexicon*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780198712053.001.0001>
- Noonan, Michael. 1999. "Non-Structuralist Syntax." In *Functionalism and Formalism in Linguistics: General Papers*, edited by Michael Darnel, 11–33. Philadelphia: John Benjamins Publishing Company.
<https://doi.org/10.1075/slcs.41.03noo>

- Orlean, Susan. 2011. "The Dog Star." *The New Yorker*. Available at: <https://www.newyorker.com/magazine/2011/08/29/the-dog-star>
- Peregrin, Jaroslav. 2014a. "Implicit Rules." *Organon F* 21 (3): 381–98.
- Peregrin, Jaroslav. 2014b. *Inferentialism: Why Rules Matter*. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137452962>
- Pollock, Joanna. 2014. "Mental Content, Holism and Communication." PhD. thesis. University of Edinburgh. Available at: <https://www.era.lib.ed.ac.uk/handle/1842/9853>
- Quine, Willard Van Orman. 1960. *Word and Object*. Cambridge: MIT Press.
- Rapaport, William J. 2000. "How to Pass a Turing Test: Syntactic Semantics, Natural Language Understanding, and First-Person Cognition." *Journal of Logic, Language, and Information* 9 (4): 467–90. <https://doi.org/10.1023/A:1008319409770>
- Rapaport, William J. 2003. "What Did You Mean by That: Misunderstanding, Negotiation, and Syntactic Semantics." *Minds and Machines* 13 (3): 397–427. <https://doi.org/10.1023/A:1024145126190>
- Recanati, Francois. 2004. *Literal Meaning*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511615382>
- Recanati, Francois. 2010. "Pragmatic Enrichment." In *Routledge Companion to Philosophy of Language*, edited by Delia Fara and Gillian Russell, 67–78. New York: Routledge.
- Shapiro, Stewart. 2006. *Vagueness in Context*. Oxford: Clarendon Press. <https://doi.org/10.1093/acprof:oso/9780199280391.001.0001>
- Travis, Charles. 1997. "Pragmatics." In *A Companion to the Philosophy of Language*, edited by Bob Hale and Crispin Wright, 87–107. Oxford: Blackwell. <https://doi.org/10.1002/9781118972090.ch6>
- Waismann, Friedrich. 1945. "Verifiability." *Proceedings of the Aristotelian Society, Supplementary Volumes* 19 (1): 119–50.
- Weiss, Bernhard. 2010. *How to Understand Language*. Durham: Acumen. <https://doi.org/10.1017/UPO9781844654468>
- Wernig, Markus, and Erica Cosentino. 2017. "The Interaction of Bayesian Pragmatics and Lexical Semantics in Linguistic Interpretation." In *Proceedings of the 39th Annual Conference of the Cognitive Science Society*, edited by Glenn Gunzelmann, Andrew Howes, Thora Tenbrink, and Eddy Davelaar, 3504–09. Austin, TX: Cognitive Science Society.
- Wilson, Mark. 1982. "Predicate Meets Property." *The Philosophical Review* 91 (4): 549–89. <https://doi.org/10.2307/2184801>
- Wilson, Mark. 2006. *Wandering Significance: An Essay on Conceptual Behavior*. Oxford: Oxford University Press. <https://doi.org/10.1093/acprof:oso/9780199269259.001.0001>

Genuinely Constitutive Rules

Bartosz Kaluziński*

Received: 19 November 2018 / Accepted: 28 January 2019


Abstract: In this article I am going to argue that despite the fact that (1) there is nothing specific to the form of constitutive rules and (2) that in some broad sense every rule has a constitutive aspect, there is a substantial difference between what might be called trivially and genuinely constitutive rules, and the difference can be spotted by looking at practices that rules are supposed to constitute, not at these rules.

Keywords: Constitutive rules; deep conventions; regulative rules; rules of efficiency.


1. Introduction

It was John Searle (1969, 1995, 2005) who popularized the notion of constitutive rules. The basic intuition that lies behind that kind of rules is best expressed by the following slogan: if there were no rules, there would be no practice. If there were no rules of chess (football, rugby, etc.), there would be no chess (football, rugby, etc.) matches. For a long time that intuition has been accompanied by a claim that constitutive rules have

* Adam Mickiewicz University in Poznań

 <https://orcid.org/0000-0001-5796-4925>

 Institute of Philosophy, Faculty of Social Sciences, Adam Mickiewicz University in Poznań Szamarzewskiego 89c, 60-568 Poznań, Poland

 bartosz.kaluzinski@amu.edu.pl



a specific form: X counts as Y in C .¹ These rules were usually opposed to regulative rules, which govern our behaviour and say what we must/should (not) do (e.g. if you are going to the opera, wear a tie). Importantly, the activities governed by regulative rules are possible to execute without any rules—it is simply possible to wear a tie even when there are no rules in regard to what should be worn on certain specific occasions.

But there are some controversies if that picture is all right. It has been argued that rules creating new forms of practices can be reduced to rules that are regulative. Suppose that we have only rules of the form Searle attributed to constitutive rules, this is “ X counts as Y in C ” (note that these rules are akin definitions that introduce new terms for types of actions/objects). The rule “such and such piece of wood counts as bishop” is basically useless without being accompanied by a set of rules that determine which moves within the game of chess are allowed, forbidden and necessary, and these rules have rather the form that was attributed to regulative rules (e.g. “In C , do $X!$ ”). If it is the case, constitutive rules are indeed reducible to regulative ones and their function is, at best, practical/mnemonic (Hindriks and Guala 2014; Guala and Hindriks 2015). According to Hindriks, constitutive rules provide us with labels of statuses established by regulative rules: “there is an underlying reality that constitutive rules serve to make apparent” (Hindriks 2009, 237).

Moreover, it has been noted (Giddens 1984) that all rules are in some sense constitutive. Even the rules of etiquette, that Searle claimed to be a paradigm example of regulative rules, are somehow constitutive. Why is that so? Because these rules are naturally in accord with basic intuition lying behind the very distinction between constitutive and regulative rules; we can justifiably claim that “if there were no rules (of etiquette), there would be no practice (of social etiquette).” If these arguments are correct, then it appears that there are indeed no constitutive rules that are *substantially* different from regulative ones. As Hindriks and Guala claim:

The distinction between regulative and constitutive rules obscures the fact that both etiquette and chess are institutional

¹ X stands for some natural phenomena (e.g. throwing the ball into the basket) and Y gives institutional description of X (e.g. scoring points), C stands for circumstances.

phenomena. A better appreciation of the relation between regulative and constitutive rules makes clear that they are cogwheels of the same social machine, even though they display different grammatical forms. (Hindriks and Guala 2014, 18)

But that conclusion, I suppose, would be far-fetched. In my brief article, I am going to argue that even if it is the case that rules of the form “ X counts as Y in C ” do not “create” practices or institutions by themselves, there is still a possibility to distinguish between a *genuinely* constitutive set of rules and a *trivially* constitutive set of rules, when we consider the background such rules operate within.

2. Layers of rule-constituted practices

Rules of the form “ X counts as Y in C ” do not make certain natural actions possible. These rules are *quasi*-definitions of institutional terms (e.g. offside, knock-down, castling etc.), and it is obvious that *physical activities* (standing in a certain place at a football pitch, punching someone so hard that they fall down, moving certain wooden pieces, etc.) that X -terms are supposed to denote can be performed without any rules. Hindriks and Guala even claimed that:

all that constitutive rules do in comparison to regulative rules is to introduce labels or names (such as ‘money’ or ‘property’) for the statuses that figure in those regulative rules. (Hindriks and Guala 2014, 19)

Indeed it appears that constitutive rules of the form “ X count as Y in C ” do not create certain activities, but they give us some names/labels that we can use when describing certain institutional activities (cf. Ruben 1997).

Arguments made by Hindriks (2009), Hindriks and Guala (2014), Guala and Hindriks (2015), and Giddens (1984) may seem good, but they miss something important. Let me start with rather an uncontroversial claim that it is not the case that there is some analogy between rules of games and constitutive rules in general, but rather if there are constitutive rules, rules of games are paradigm examples of them. So, taking a look on rules of games is probably the best way to acquire some knowledge concerning

constitutive rules. The rulebook of chess (or any other game) contains both rules that are definitions or specifications and rules that determine which moves are allowed, necessary and forbidden within the game (for example, you cannot move the rook diagonally; the knight can be moved to a square that is two squares away horizontally and one square vertically, or two squares vertically and one square horizontally; the game is started by white). Thus, the game is constituted by the set of rules that specifies certain things (for instance, the shape of the figures) and determines what must/can (not) be done within that game.

But when proponents of the thesis that the difference between constitutive and regulative rules is a merely linguistic focus their attention on the form of rules, they miss that the practice of playing, for instance, chess, cannot be established *merely* by rules of that practice (Schwyzer 1969; Marmor 2007). Why is that so? What, besides rules, do we actually need to play a game? To answer those questions, let me start with a reminder of a thought experiment made by Schwyzer (1969). Suppose that you are watching two people from the other side of the globe. They do not violate rules of chess (they move pieces as they are allowed to), but their moves are rather chaotic and there is no way to spot what strategy they have adopted, why they are making these moves rather than others. After some time of this disorganized “game” there happens to be a checkmate—and now one side is deeply terrified and the other one deeply relieved. In that scenario, these people did not *play* chess, but they rather participated in some sort of *ritual* that is to determine, for example, members of which tribe will suffer from plagues sent down by gods. For it is possible that two people move wooden pieces in accordance with the rules of chess (pawns one square forward, bishops diagonally, etc.), but they do not *play* chess. They simply ascribe different socio-cultural sense to actions they undertake—they treat their activity as a religious ritual rather than a competitive game.

Hence, it is not sufficient to act in accordance with the rules of the practice to participate in *that* practice. Something that might be called “deep convention” (Marmor 2007; cf. Roversi 2014) should also be known by the participants. In case of chess, it would be a convention of *playing a competitive game* which, roughly speaking, may include knowledge that:

1. Games have objectives, and the ultimate objective of a game is to *win*.
2. Games are “detached” from ordinary life (for instance, when a rugby player knocks down a member of the opposite team, they are not subject to criminal charges concerning assault).
3. It is possible to distinguish participants in the practice and non-participants, spectators (Marmor 2007).

But it is not the whole story. There are deep conventions that underpin rule-constituted practices,² but also there is something that emerges upon the rules of the practice. One more time, let’s get back to chess. There are many different types of openings, attacks and defenses—they have fancy names,³ and they specify what moves in certain circumstances should be made to achieve certain aim. Note that these rules (they might be called “rules of efficiency”) are not identical to constitutive rules of chess nor are they mere paraphrases of them. In fact, it is rather the case that one can play chess without following them or even being aware of their existence. These rules are instrumental; they determine what to do within the framework of the game to eventually win the game (whereas the concept of winning is part of the deep convention of playing the competitive game). Moreover, sometimes these rules of efficiency tell us to violate some rules of the game. It is rather not possible in the case of chess, but when we play, for instance, basketball it is quite common that the losing team, when the end of the game is quite near, start to foul members of the opposite team (especially players that are bad at free throws), hoping that they will miss free throws and the losing team will have an opportunity to start their own action immediately after the miss. As Roversi noted, these rules are distinct

² As Roversi noted, “Constitutive rules alone cannot give us genuine reasons for action unless they are embedded in a context already endowed with a social meaning: there would be no point in following the rules of chess if chess were not a game or another sort of social activity, just as it would be meaningless to realize all the formal conditions for transferring property if these conditions were not part of a legal practice” (Roversi 2014, 210).

³ There is a big Wikipedia entry that enlists them: https://en.wikipedia.org/wiki/List_of_chess_openings

from constitutive rules but “they denote situations that can be realized only by instantiating a given game’s institutional, rule-constituted elements” (Roversi 2014, 212). This is, any recommendation of as how to checkmate your opponent cannot be made before the rules concerning checkmate are in force.

It appears that, in case of such practices as games, rules that are listed in official rulebooks are not sufficient to create them. It is not possible to participate in any *game* without knowing the deep convention of playing a competitive game. There is another set of rules that emerge upon such practices as games: rules of efficiency. It is obviously possible to participate in a practice without knowing or adhering to these rules (probably children and, more generally, beginners and amateurs do not know these rules), but such games would be gawky and, given the ultimate objective of the practice (winning), the emergence of such rules is perfectly understandable. Hence, there is no metaphysical necessity in following these rules to participate in the practice, but *normally* we learn and use them because we aim at achieving the ultimate goal of the practice (in case of competitive games it is winning).

3. Games and etiquette

Now take a look at etiquette, which has been for a long time, treated as the best example of a practice which rules are merely regulative ones. However, it was noted that if there were no rules of etiquette there would be no social practice of etiquette, so the rules of etiquette seem to fit well into the basic intuition lying behind the very distinction of constitutive rules. But there is an important difference between games and etiquette. Namely, there are other requirements that are put on a participant of the practice. To play a game, for instance chess, you have to:

1. Know the deep convention of playing chess and intend to play.
2. Know at least at some working level,⁴ rules of chess.

⁴ This is to say, that players do not need to recite relevant passage from the rulebook. Nevertheless, it would be really odd to claim that one can play a game without any knowledge of its rules.

But to play well, you also need to:

3. Know the rules of efficiency (and use them properly).

So, there are many things that one needs to know and intend to do to participate in a game (cf. Kaluziński 2018a). And things seem to be much different in case of etiquette. There are rules that we should eat using a fork and knife (and not with our bare hands), wear a tie when going to the opera, etc. and it looks that it is all that we need to know to participate in such a practice as social etiquette.⁵ Perhaps I am missing something, but it appears to me that there is no deep convention underpinning etiquette (there is nothing similar to the concept of winning that needs to be known to participate in a practice of playing chess or rugby). It is also the case that a person can participate in a practice of etiquette by merely acting in accordance with its rules (for instance, wearing a tie when going to the opera). And, lastly, there seems to be no rules of efficiency that determine how to “be good at etiquette,” while there definitely are rules that tell us how to be good at chess, football and rugby. If there is any sense in speaking of “being good at etiquette,” it simply consists in following its rules.

Summarizing, practices that intuitively are constituted by systems of rules (for instance, games) are indeed quite complicated. These practices are underpinned by deep conventions (in case of games, the convention of playing a competitive game). There are also rules of efficiency that provide us with the means of achieving the ultimate goal of the practice (e.g., winning). It appears that in case of such practices as etiquette, which was considered by Searle (1969) to be a practice that is not constituted by rules, things are very different and there is only one “layer of practice”—rules of etiquette.

But perhaps things are little different and there is no single “practice of social etiquette,” but there are many social practices that are governed by

⁵ One may wonder if it is not the case that we would participate in the practice of social etiquette even in the case if we did not know its rules, because if we breach some rule of etiquette (e.g., we would try to eat soup by dinking it directly from the pot), we would face disbelief and critique from members of our society regardless of our knowledge of that rule. If it is true, the difference between participation in practice of etiquette and in such practices as games is even bigger.

the rules of etiquette. In such a case, one may argue, it is possible to find some deep conventions (e.g. *being in the boss/subordinate relationship*) and my argument that we can identify/specify constitutive rules in terms of the broader practices in which those rules operate within is flawed. I have certain doubts concerning such an account. It seems analogous to the argument that there is no such thing as games and rules of games in general but only rules of specific games: chess, bridge, ice hockey, rugby etc. One may find appealing to such an account unattractive. But, for the sake of argument, let me assume that it is correct. Does it pose a grave challenge to my account? I tend to think that it does not. Of course, if one looks carefully at certain social practices, then one can spot various deep conventions. But none of them is a deep convention that *underpins specifically* etiquette. It might be the case that the rules of etiquette tell us how subordinates should behave towards their boss and vice versa but the very concept of business hierarchy (or chain of command) does not pertain to deep convention of etiquette but rather to the deep convention of business corporation (or military).

It appears that there is no deep convention that underpins etiquette as such but there are deep conventions that are cornerstones for various social practices or institutions like corporations or the army and our behaviour within such practices or institutions can be guided by rules of etiquette. But obviously these rules of etiquette are not realizations of deep conventions in the same sense as the rules of rugby are a realization of the deep convention of playing competitive games. Deep conventions, to use Marmor's words, are "enabling the emergence of some of the surface conventions that we normally follow" (Marmor 2007, 586) and clearly business hierarchies or chain of commands do not play that role for the rules of etiquette. Suppose that a boss in a certain enterprise is extremely polite because she wants her employees to feel valued and highly motivated, so they will work efficiently. Of course, in such circumstances the boss is following the rules of etiquette but she is doing so because the deep convention of business corporation includes the aim of maximizing profits and she believes that being polite is a good way of increasing her employees productivity. Once again, there are some deep conventions and rules of efficiency included in the reasoning that the boss makes but they do not pertain to the practice of etiquette but rather to business corporation. Note that it is not

uncommon that in certain circumstances different deep conventions can be spotted, if we look carefully enough. For instance, when playing a professional football match, players know two deep conventions: 1) convention of playing competitive games that is necessary for their participation in that practice and 2) a convention of “chain of command” or “boss/subordinate hierarchy” that determines that they follow tactical instructions made by their team manager rather than chaotically run around the pitch. It is important not to conflate those deep conventions: only the knowledge of deep convention of playing competitive game is a necessary condition for a football match to occur, while the convention of “chain of command” is *incidental* for the possibility of participation in a match.

Perhaps that objection could be phrased differently. Imagine a scenario in which the deep convention is “making a workplace a more comfortable place for everyone” (without any devious intentions). This deep convention pertains to the practice of “workplace etiquette” and the rules of workplace etiquette open the space for the rules of efficiency. Clearly, there are no deep conventions that pertain to etiquette as such, but probably we could find deep conventions that pertain to workplace etiquette (and many other types of etiquette). That is interesting point but I think that such an argument is problematic. It appears that one can justifiably claim that such account is redundant because the deep convention of “making a workplace a more comfortable place for everyone” probably does not introduce new concepts (like winning in case of playing competitive games) nor it set some new requirements on participants in the practice. One can participate in the practice of workplace etiquette without any knowledge of such deep convention (whereas one cannot participate in the practice of playing rugby without knowledge of the convention of playing competitive games). Perhaps I am missing something, but it appears to me that there is no point in postulating such deep convention as “making a workplace a more comfortable place for everyone.”

Lastly, one might wonder if the existence of rules of efficiency is really characteristic only for such practices as games and there is no room for such kind of rules in the case of etiquette. After all, one can use the rules of etiquette in such a way that she achieves her aims, for instance when one ignores greeting from person S and by doing so causes S to understand that

S is an unwelcome guest. But even if it is the case that there are some rules that emerge on the basis of the rules of etiquette, are they the same kind of rules as rules of efficiency? I tend to think that they are not. It is crucial to remember that the deep convention that underpins rule-constituted practice can introduce new concepts and determine the ultimate aim of that practice (for instance, winning). Rules of efficiency tell us what to do in order to achieve these ultimate aims (how to win a chess match; what to do in order to win a war or scare off potential enemies, etc.) There is no similar thing in the case of etiquette, because there is no deep convention that underpins *specifically* etiquette. So, even if there is something like the rules of efficiency in the case of etiquette, these rules lack the feature of giving us recommendations as to how to achieve the ultimate aims of etiquette simply because there is no deep convention that specifies what the ultimate aim of etiquette is.

One might wonder if indeed there are no deep conventions that underpin certain etiquette regulated behaviour.⁶ But I am not sure if we can spot a deep convention that is broader and more informative than “acting appropriately,” “exercising proper behaviour” or similar. Treating “exercising appropriate behaviour” as a deep convention of etiquette raises important question: if such “deep convention” is able to put similar requirements on participants in the practice of etiquette as deep convention of playing competitive game does? Going back to the thought experiment conducted by Schwyzer, one needs to ask if it is possible to participate in practice of etiquette without the knowledge of such “deep convention?” In the case of games it is clear that knowledge of deep convention is a necessary condition of playing chess or basketball, but I am sceptical if it is the case that we cannot for instance greet our colleague/boss by saying “Good morning” without the knowledge of anything more than the knowledge of the etiquette rule that tells us that we should greet our colleague/boss by saying “Good morning.”

⁶ Marmor would be reluctant to accept that claim (Marmor 2007, 606).

4. Genuinely and trivially constitutive rules

Finally, I would like to introduce a distinction between *genuinely constitutive rules* and *trivially constitutive rules*. Rules of the former kind create what might be called “multi-layer practices” (for instance, games), whether rules of the latter kind create “one-layer practices” (for instance, etiquette). As it was noted in the previous sections, games are not just systems of constitutive rules, as Searle thought, but they are three-layer practices, which consist of deep convention, the system of rules, and rules of efficiency. It appears that social practice of etiquette is indeed in some very broad sense constituted by rules, but these rules are not underpinned by deep conventions and there are no rules of efficiency built upon rules of etiquette. So, we can uphold the claim that in a broad sense all rules might have a constitutive aspect, and it is right to say that “if there were no rules of etiquette, there would be no practice of etiquette.” Probably, it is possible to justifiably claim something like that about virtually any rule-involving practice. But I am convinced that we are not doomed to the claim that all rules are constitutive in the exact same (and blatantly trivial) sense. Some rules are trivially constitutive and some are genuinely constitutive. The whole point is not to look at rules (because indeed there is nothing special about the form of the rules of either kind) but to look at the things that are supposed to be constituted by rules, i.e., practices or institutions. What is really important is what the effect of the existence of a certain set of rules is, not what form these rules have. And that effect could be multi-layer or one-layer practice.

It appears to me that, when discussing constitutive rules, too much attention has been paid to the form of constitutive rules, and too little to the things they are supposed to constitute. Analogically, it is like in the case when we tried to investigate what the *means of production* are by focusing on their physical features (e.g., that such and such machine is made of steel) and fail to notice if it can actually produce anything (and if it is able to produce something more than noise when working).

Lastly, one may wonder if the considerations presented above can be extrapolated from games to other practices. Take, for example, such practice as the legislative process. It appears that there is some deep convention

lying behind it; namely, concepts of authority or representation are needed (for instance, we should know that members of parliament act not as private persons but as legitimate representatives of the society; one that does not have concepts of authority or representation would be clueless when watching a broadcast of parliamentary debates). Then there is a “proper” set of rules (enlisted in constitution and some other legal acts) that determine how new laws are passed: who can be a member of parliament (e.g., age census), how MPs are elected (electoral law), and how decisions in the parliament are to be made (e.g., by majority of votes in the presence of at least half the number of members of parliament and then the bill should be signed by the president). And finally, there is a set of rules of efficiency that tell us how to use powers and duties introduced by these “proper” rules to achieve ultimate goal of a practice (passing the bill smoothly or blocking the bill by using, for instance, filibuster). Hence, at the first glance, the analysis provided in this brief paper is not limited to such practices as games.

There is one important clarification I should make at this point. I claim that genuinely constitutive rules take part in creating multi-layer practices and one may wonder if it is not the case that we can make any rules genuinely constitutive ones by just supplementing trivially constitutive rules (for instance rules of etiquette) with two additional layers, e.g. by (A) a rule that everybody should strive to be better at etiquette than everybody else and (B) creating rules on how to effectively achieve this. When I say that genuinely constitutive rules take part in the creation of multi-layer practice I mean that there is a “cascade” of layers. First, we have deep convention that underpins a practice (for instance, the convention of playing a competitive game). Without the knowledge of deep convention there is no possibility to participate in the relevant practice. Second, there is a system of constitutive rules (the rulebook of chess, rugby, football, etc.). Rules of this kind are just different forms of the realization of the very general idea of competitive games. Third, there are rules of efficiency that tell us how to achieve our aims (and, most importantly, the ultimate aim of the practice) within the frame of the practice that is made by the system of rules. Hence, one cannot turn rules of etiquette into genuinely constitutive rules just by adding arbitrarily two kinds of rules because these rules do not meet the criteria deep conventions and rules of efficiency need to meet: they are not

necessary for the very possibility of participation in a practice; rules of etiquette do not seem to be a form of realization of (A) like the rules of basketball are a form of realization of the deep convention of playing competitive games; since there is no ultimate aim of the practice of etiquette there are no rules of efficiency that recommend to us certain moves that may help us achieve that aim.

Even if it is the case that traditionally understood constitutive rules (as having the form “ X count as Y in C ”) are reducible to rules that have form attributed to regulative rules, there is still a possibility to differentiate *genuinely* constitutive rules from *trivially* constitutive ones. We just need to not look at the form of the rules, but rather at the reality that they are supposed to “create.” If that part of reality is a simple, one-layer practice, then rules that “constitute” that practice are constitutive only in a broad and trivial sense. We can indeed say that “if there were no rules, there would be no practice.” However, this can probably be said about all rule-involving practices. But there are also rules that are part of multi-layer practices or institutions. These rules are genuinely and non-trivially constitutive. Why? Because they take part in creation of complicated practices that pose other requirements for participants in these practices. To participate in an etiquette-like practice, it is sufficient that a person follows (or not) its rules or, perhaps, it is even the case that she can participate in it without awareness of these rules. But to participate in, for instance, a game of chess one needs to know *relevant* deep convention, intend to play and being committed to following (at least some subset of) its rules. Normally, one also needs to know and properly use rules of efficiency to have a shot at winning (achieving the ultimate goal of the practice of playing a game).

Summarising, even if Guala and Hindriks (2015) or Giddens (1984) are right to claim that rules that traditionally have been called “constitutive” are reducible to regulative rules, and, then, even paradigm example of regulative rules appears to fit into the slogan “if there were no rules, there would be no practice,” it is not the case that we cannot distinguish genuinely and trivially constitutive rules. Genuinely constitutive rules are part of multi-layer practices. To spot them, we need to look not at rules, but rather at practices that these rules are supposed to constitute. If we look carefully enough, we can spot if there is a deep convention underpinning

the practice, and if there are rules of efficiency built upon these rules. That might be hard, but it is necessary to spot non-trivially constitutive rules. Hence, when Guala and Hindriks say that “constitutive rules are, at roots, just regulative rules dressed up in institutional language” (Guala and Hindriks 2015, 189), they are mistaken if their claim is to be understood as applied to the function of rules. Perhaps the slogan that expresses the basic intuition lying behind the notion of *genuinely constitutive rules* should be “If there were no rules and deep conventions (plus, normally, rules of efficiency), then would be no practice.”

5. Conclusions

In the most orthodox account of constitutive rules they are characterized by their form and opposed to regulative rules, but there are also alternative views that emphasize that constitutive rules have a strong normative aspect of determining what should (not) or may be done within the practice (cf. Ransdell 1973; Hindriks 2009; Hindriks and Guala 2014; Guala and Hindriks 2015; Kaluziński 2018b). These accounts face a challenge: they need to provide some new characterization of constitutive rules that do not blur the distinction of constitutive rules from other types of rules. Most notable accounts of Hindriks (2009), Hindriks and Guala (2014), Guala and Hindriks (2015) steer in the direction of “constitutive rules reductionism:” they claim that constitutive rules are reducible to regulative ones and their sole function is to provide us with labels that are referring to statuses established by regulative rules. Their role is purely practical/mnemonic. In this paper I tried to show that this is not the only viable option. I argue that there is another way to specify constitutive rules in terms of the broader practices in which those rules are applied. If we take deep conventions and rules of efficiency into account, we can identify genuinely constitutive rules even if we emphasize the normative side of these rules.

Funding

The work on this paper was funded by National Science Center, Poland, grant under award number 2015/19/D/HS1/00968.

Acknowledgements

I would like to thank two anonymous reviewers of this journal whose insightful comments helped me improve my paper.

References

- Giddens, Anthony. 1984. *The Constitution of Society*. Berkeley: University of California Press.
- Guala, Francesco, and Hindriks, Frank. 2015. "A Unified Social Ontology." *The Philosophical Quarterly* 65 (259): 178–201. <https://doi.org/10.1093/pq/pqu072>
- Hindriks, Frank. 2009. "Constitutive Rules, Language, and Ontology." *Erkenntnis* 71 (2): 253–75. <https://doi.org/10.1007/s10670-009-9178-6>
- Hindriks, Frank, and Guala, Francesco. (2014), "Institutions, Rules, and Equilibria: A Unified Theory." *Journal of Institutional Economics* 11 (3): 459–80. <https://doi.org/10.1017/S1744137414000496>
- Kaluziński, Bartosz. 2018a. "What Does It Mean that Constitutive Rules Are in Force?" *Argumenta* 4 (1): 111–12. <https://doi.org/10.14275/2465-2334/20187.kal>
- Kaluziński, Bartosz. 2018b. "Rules and Games." *Philosophia*. <https://doi.org/10.1007/s11406-018-0050-2>
- Marmor, Andrei. 2007. "Deep Conventions." *Philosophy and Phenomenological Research* 74 (3): 586–610. <https://doi.org/10.1111/j.1933-1592.2007.00041.x>
- Ransdell, Joseph. 1971. "Constitutive Rules and Speech-Act Analysis." *Journal of Philosophy* 68 (13): 385–99. <https://doi.org/10.2307/2025037>
- Roversi, Corrado. 2014. "Conceptualizing Institutions." *Phenomenology and the Cognitive Sciences* 13 (1): 201–15. <https://doi.org/10.1007/s11097-013-9326-y>
- Ruben, David-Hillel. 1997. "John Searle's *The Construction of Social Reality*." *Philosophy and Phenomenological Research* 57 (2): 443–47. <https://doi.org/10.2307/2953734>
- Schwyzler, Hubert. 1969. "Rules and Practices." *The Philosophical Review* 78 (4): 451–67. <https://doi.org/10.2307/2184198>
- Searle, John. 1969. *Speech Acts: An Essay in the Philosophy of Language*. Oxford: Oxford. <https://doi.org/10.1017/CBO9781139173438>
- Searle, John. 1995. *The Construction of Social Reality*. New York: The Free Press.
- Searle, John. 2005. "What Is an Institution?" *Journal of Institutional Economics*, 1 (1): 1–22. <https://doi.org/10.1017/S1744137405000020>

The Morality of Euthanasia

Adam Greif*


Received: 22 November 2018 / Accepted: 16 March 2019


Abstract: In this paper, I defend the view that the requested euthanasia of adults is morally permissible and should be legalised; I use an argument from analogy which compares physician-assisted euthanasia with morally less ambiguous and, in my opinion, an acceptable instance of mercy killing. I also respond to several objections that either try to prove that the instance of mercy killing is not acceptable, or that there is a fundamental difference between these two cases of killing. Furthermore, in the remainder of the paper I defend the moral permissibility and legalisation of euthanasia against several objections that appeared in local disputes on this issue, based on the concepts of the limits of freedom, the slippery slope, and the needlessness of euthanasia.

Keywords: Legalisation; moral freedom; morality; needlessness of euthanasia; requested euthanasia; sanctity; slippery slope.

There are two types of death: we either die at a time and in a manner we do not choose or at a time and in a manner which we do. We can state that most people who have died did not depart at the time or in the manner they would have chosen, as they died either earlier or in a different way than they wanted to. We can assume that, if given the chance, those who

* Comenius University

 Department of Philosophy and History of Philosophy, Faculty of Arts, Comenius University, Šafárikovo námestie 6, 814 99 Bratislava, Slovak Republic

 greif2@uniba.sk



died from illnesses, injuries, advanced age, unfortunate accidents, or by being killed would have preferred to go at a later time as well as in a different and less unpleasant way. In light of this idea, it seems preferable to meet death on our own terms. If medicine and technology have made this possible, we should probably want to allow people to be in charge of the circumstances of their own death.

This opinion does not invite controversy when talking about people who wish to live longer, but it becomes controversial once we start discussing those who want to shorten their lives. Granting a longer life to people who desire it does not appear to be an issue, so why does it become one once we start contemplating the rightness of hastening the death of those who request it?

Some opine that we would commit a serious mistake by bringing death to those who ask for it, arguing that obeying such a wish is unreasonable. While some claim that wishing to die is unreasonable because every life is worth living, others claim that it is impossible to rationally conclude that our life has ceased to be worth living. Critics of euthanasia thus suggest that rather than complying with a patient's wish to die, we should strive to improve their life.

These objections are as important as they are philosophically interesting, and a decent defence of euthanasia should be able to respond to them. Although I share the opinion that we should primarily endeavour to improve the lives of those wishing to die and that a person can sometimes err in thinking their life has ceased to be worth living, I also maintain that in certain circumstances a person's life might become not worth living and that they can rationally arrive at such a conclusion. However, I defended these opinions elsewhere and therefore will simply assume their truth in this paper (see Greif 2018). If they were not assumed true, there would be little need to respond to the second type of objections that question the morality of ending another person's life prematurely, i.e., euthanasia.

Euthanasia poses an ethical dilemma. On one hand, there are people who do not consider their lives worth living and express an honest wish to die. I assume we want to empathise with their suffering and show respect for their decision to end their life. On the other hand, a moral doubt remains lingering: "Is it right to end someone's life on their request? Would we not

be committing murder or another serious offence?” And we surely do not want to err in such a serious matter. How can it be resolved then?

I think there is a solution to this problem. To demonstrate it, I will first present an argument in favour of requested euthanasia being morally unproblematic and permissible.¹ I will support the argument by responding to several objections to its premises. Then I will focus on three objections that appeared in a debate in the Czech journal *Filosofický časopis* in 2010 through 2012. I will try to prove that all these objections can be countered. My defence only focuses on the requested (or voluntary) euthanasia of adults, which I define as “an act when an adult person kills another adult person for the latter’s well-being and on the latter’s request.”² My main objective is to defend the thesis that “the requested euthanasia of adults is morally permissible.”

Before commencing my defence of euthanasia, I would like to comment on the nature of my defence. Ethics is a delicate discipline, and ethical questions are notoriously difficult to resolve in a way that everyone would find satisfactory. As a consequence, a certain tension characterises my defence, which I would prefer to disclose at the very beginning. I understand ethics as a kind of rational and secular debate. Therefore, “ethics” deferring to a religious authority or an authority of individuals who claim to possess special moral knowledge is not what I consider to be ethics. I find it crucial that in ethics, whenever possible and necessary, moral assertions should be justified without referring to personal or divine authority. On the other hand, justifying every statement is neither necessary nor possible. Some key statements I cannot or will not justify, hoping that they do not require justification. I am referring to statements that philosophy would call “moral intuitions.” My judgement about the trapped lorry driver case, which I will present shortly, provides a rather illustrative example of moral intuition.

¹ I use the expressions “morally unproblematic,” “permissible,” “permitted,” “right,” and “moral” interchangeably. Correspondingly, “morally prohibited,” “not morally permissible,” “wrong,” and “immoral” are also interchangeable.

² I base this definition on the etymology of the word “euthanasia”—*εὐθανασία* (euthanasía), from *εὖ-* (eu, “well” or “good”) + *θάνατος* (thanatos; “death”)—that is, good death. Since I want to differentiate euthanasia from suicide, I define it as good death caused by another person.

This judgement plays a crucial role in my defence of euthanasia which would not work without it. It is, however, unjustified, and I do not know with certainty if the reasonable majority would accept it. Taking into consideration how much disagreement moral questions tend to generate, it is possible that even what I consider utterly evident might invite harsh criticism. Although I do not know if the reasonable majority would agree with me, I will assume it does. Naturally, in cases like this claims can be made that mostly suit the author's interests instead of reflecting the reasonable majority's opinion; however, this risk is present in every ethical debate and should be accepted by its participants. If we are to arrive at some kind of a resolution, we, the disputing sides, have to establish an initial agreement, even if only by guessing.

* * *

Let us advance to the argument put forth in favour of the morality of euthanasia. It is based on the famous "trapped lorry driver case," in which a man is killed in a way that I consider permissible and that shares every essential feature with physician-assisted requested euthanasia.

The trapped lorry driver case

A driver is trapped in a blazing lorry. There is no way in which he can be saved. He will soon burn to death. A friend of the driver is standing by the lorry. This friend has a gun and is a good shot. The driver asks this friend to shoot him dead. It will be less painful for him to be shot than to burn to death. Should the friend shoot the driver dead? (Hope et al. 2008, 185)

Premise 1: I can only speak for myself with certainty, but I cannot think of any moral reason why the friend should not comply with the driver's request. Therefore, I believe that *the friend should shoot the driver dead*. However, before killing the driver, I would have to verify that he truly cannot be helped and that he would suffer tremendously if I did not shoot him. But if I was certain that his situation really was awful and desperate, I hope I could muster enough courage and presence of mind to shoot him dead. I would consider it the right thing to do.

In my opinion, emotions tend to get tangled up with morals when discussing euthanasia, which in turn might cloud our judgement. I am sure

that shooting the driver dead would not make me happy. I would wish that I did not have to do it and that the lorry driver could live. Because of these feelings of unhappiness and tragedy, I might be tempted to think that I did something wrong. But my reason would clearly tell me that it is not so, as the only possible alternative would have been much worse. If I did not kill him, the trucker would have suffered more and would not have gained anything. So if I were to find myself in such a situation, I would act partially out of empathy, as I would not want the trucker to suffer, and partially out of respect for his free will, because I would want to comply with his wish.

I assume that most people would feel and think in a similar—if not the same—way. There are only two options: either the friend shoots the driver dead or he does not do anything. But his not doing anything is clearly worse for the driver, because he will needlessly suffer if he is not shot dead. On the other hand, even though shooting the driver would be tragic, it would still be preferable to letting him burn to death. Therefore, I do not think that any convincing objections can be raised against shooting the lorry driver dead.

Consequently, I firmly believe that shooting the driver dead in order to spare him a painful death is permissible. Nevertheless, an anonymous reviewer remarked that the argument I am about to present is unconvincing, since the first premise is unjustified. I have to admit that this is true; I provided no justification for it, as I cannot prove it and cannot support it with any further arguments. But I have so much certainty in it that if someone expressed disagreement or doubt in relation to it, I would have to ask them for an explanation. In other words, what reason could one have to think that shooting the driver dead is not permissible?

There are several objections we would normally raise against shooting someone dead. Shooting the driver dead could be wrong because (1) it is against his will; (2) deprives him of all the good things he might have experienced if he lived; (3) harms him; or because (4) it violates his right to life.

We can quickly determine that these four initial objections do not apply in the lorry driver's case. First, the lorry driver wanted to die. Second, even though it would be true for the majority of people that their death would deprive them of all the good they could have experienced if they stayed

alive, it is not true for the lorry driver, as nothing good awaited him. Third, it is true that killing a human being harms them in a way, since death typically implies bodily damage of some kind. However, we often choose to endure great bodily harm if it is in our interest. For example, we endure it during a common medical intervention, or, to mention an extreme case, as Aron Ralston endured it when he amputated his forearm with a pocketknife in order to save his life. Fourth, without any doubts, the right to life is one of the most fundamental moral principles we have. Still, we are able to renounce our rights. The lorry driver, just like a patient requesting euthanasia, had a right to life, but he renounced it because it was in his interest.

If any objections were raised against the argument I advance, I suspect that the first premise would be more likely to provoke them; therefore, I presume it needs more support.

One further objection could be that the friend should not shoot the driver dead, because there is always some alternative. In the trapped lorry driver's case, it is only *ex hypothesi* true that he cannot be saved. But in real life we will not have this certainty and there will always be a chance, however small, that the driver could be saved.

Although our intuitions may betray us in some thought experiments, I do not think this is the case now. The original version of the trapped lorry driver case, as described by R.M. Hare (1975) and pointed out by Tomáš Hříbek (2010), illustrates this well. Rather than being a hypothetical case, it describes a situation that actually happened, very similar to the one described above, and which was reported by the press at the time (the lorry driver was probably killed in the end). Instead of pondering on what we would do in a fabricated thought experiment, we can ask ourselves how we would behave if we found ourselves in such a situation, surrounded by all its uncertainties. However small, there is always a chance that the driver can be saved. But if this chance is minimal, it would be more reasonable not to risk the much more obvious possibility of the trucker burning to death coming true.

A patient requesting euthanasia is in a similar situation to the driver of the blazing lorry. Although he may not have any epistemic certainty of the accuracy of his prognosis, the evidence at hand and the medical knowledge

may yield enough practical certainty for him to make up his mind. (To read more about this type of objection, see Greif 2018).

Still another objection might be that shooting the driver dead is wrong because killing an *innocent* person is always wrong, regardless of the consequences. Given that the driver is innocent, it is wrong to kill him.

It is very difficult to deny that killing an innocent person is wrong. After all, it is one of the most fundamental moral convictions. For example, killing a random passer-by is without any doubt a heinous act. In spite of this, I still maintain that killing an innocent person differs from committing a morally prohibited killing, i.e., murder. I think the rule “killing an innocent person is wrong” is usually sufficient for us to understand what makes killing a person wrong; however, I do find it imprecise, as I believe that the trapped lorry driver’s case presents a perfect counter-example. Should we refuse to shoot the driver because of his innocence?³ What does the driver’s innocence or guiltiness have to do with the moral evaluation of his case?

I tried to show that it is morally permissible to shoot the lorry driver dead. If this judgement is right, then the rule that “killing an innocent person is wrong” cannot always be correct, as the lorry driver was also innocent. However, since this a rather intuitive argument, it should be accompanied with an explanation of why the rule is incorrect despite its undeniable appeal. In order to find it, I think we have to think about the purpose of guilt.

Why is it sometimes permissible to kill a guilty person but not an innocent one? It seems that the moment an attacker assaults someone, they temporarily lose some of their rights as we stop taking their interests into full consideration. It seems that by endangering someone else’s life, for instance, they can lose some of their rights, and their guiltiness gives us a moral right to treat them differently than we normally would. When someone commits a greater crime—say, threatens the life of an adult or a child—it might give us the right to take their life in self-defence. However, a patient requesting euthanasia has not committed any crime. We have no

³ One could even argue that killing the lorry driver is not only morally permissible but also obligatory; if we could comply with the trucker’s request but failed to do so, we would allow needless suffering to take place. However, I will not be defending this opinion here.

reason to treat them any differently than anyone else. Therefore, we should neither deny them their rights nor ignore their will.

By analysing the purpose of guilt, we can explain why the rule of “killing an innocent person is wrong” should be considered imprecise despite its intuitiveness. In almost every case when an innocent person is killed, we are entitled to think that a morally prohibited killing—that is, murder—occurred. We have a good reason to think so, because the only good argument in favour of the opposite—that the “victim” wanted to die and the killing was merciful—typically does not apply. It is rare that outside of the context of medicine or war perhaps a person was killed because he or she actually wanted to die and being killed was good for them.

Finally, some might argue that it is wrong to shoot the driver dead because his life has an *intrinsic and impersonal value*, that some would call *sanctity* (Dworkin 1993), and it is wrong to sacrifice it for the quality of one’s life or for any other value. Since the value of life itself is impersonal, it is wrong to end driver’s life, even though it is no longer valuable for him or anyone else.

There is surely much to be said about this view, but I would like to propose a somewhat minimalistic refutation. To start, I think we should ask the following question; “Is the value of life itself absolute?” In other words, does the intrinsic value of life always come before other values or only in some cases? Let us consider both options, one by one.

If the intrinsic value of life always comes before other values, then no value can ever be more important. Compared with any other value, the intrinsic value of life would have an absolute worth. If those proclaiming that “every human life is sacred” accepted this option, it would presumably lead to some consequences that would be difficult to accept. Steven Luper illustrated these outcomes by means of the following thought experiments:

Two Spells

I know how to cast two magical spells. One of them, which Mary wants me to use on her, would ensure that she has a life that is extremely good and far better than the life she otherwise would have had, but the spell will also kill her painlessly in her sleep one day sooner than the day she otherwise would have died of old age. The other spell, which she has forbidden

me to use, would not kill her but would ensure that she has a life that is wretched, and far worse than the life she otherwise would have had. (Luper 2009, 186)

The first spell would shorten Mary's life by one day; in exchange, it would ensure that the rest of her life would be extremely good. If we accepted the view that human life has an intrinsic value that can never be sacrificed for any other value, then casting the first spell would be wrong, as Mary's life would be sacrificed for her quality of life. In addition, it would be morally more wrong to apply the first spell than the second one. However, we would probably disagree with these judgements. Besides being beneficial for Mary, the first spell is perfectly moral and in fact more moral than the second one. Therefore, the view that human life has intrinsic value has some hard to accept consequences.

Let us now consider Luper's second thought experiment:

Unintentional Suicide

I have an illness that will kill me within a week if allowed to progress. There is a treatment that will extend my life by one more year, but I will be in pain nearly the entire time. I weigh the extra time against the pain involved and decide to refuse the treatment. I die three days later. (Luper 2009, 187)

In this case, I sacrificed a longer life in favour of avoiding pain. If life has absolute intrinsic value, then this would be an immoral act as it is more valuable to live longer than to avoid pain. However, this is not how we think. Refusing treatment is a perfectly rational and moral decision.

If we believe that the intrinsic value of life always comes before other values, it leads to certain consequences that we are reluctant to accept. This is why I do not think that we should accept the strong version of the principle that every human life has intrinsic value.

Let us consider the second, weaker version of the principle, according to which the intrinsic value of life only comes before other values in some cases. If the intrinsic value of life only comes first in certain situations, then sometimes are other values, such as the quality of life or personal dignity, more important. So when a patient requests euthanasia, how do we know whether these values did not outweigh the value of their life being

intrinsically valuable? Considering the Luper's thought experiments, it seems fairly common that the considerations of quality of one's life or of one's dignity are more important than the intrinsic value of life.

Premise 2: If the friend killed the lorry driver, it would be a case of requested euthanasia, as the driver would die for his own good and on his own request. My second premise is that *shooting the lorry driver is analogous to physician-assisted requested euthanasia*. When calling it euthanasia, it does not matter who performs it; the only fact that counts is that person A kills person B for person B's benefit and on person B's request. These conditions are met in the trapped lorry driver's case. We can furthermore assert that the friend would shoot the driver dead with the intention of helping his friend and complying with his wish rather than taking revenge on him or wanting to put an end to his annoying screams.

So when a physician administers the lethal shot to a patient, for the patient's benefit, on the patient's request, and motivated by a good intention, his act is considered right because he does what the friend would do to the trapped lorry driver by shooting him dead. Thus, the argument goes as follows:

Argument from analogy

P1 Shooting the driver is morally permissible.

P2 Shooting the driver is analogous to physician-assisted requested euthanasia (assuming that the physician performs the act with good intention).

C Physician-assisted requested medical euthanasia is presumed to be morally permissible (assuming that the physician performs the act with good intention).

Any rejection of the second premise implies that there is a difference between the two instances of killing. This is true; we could surely find plenty of differences between killing the driver of a blazing lorry and killing a particular patient. However, not all of these differences are relevant. For example, in the trapped lorry driver's case, there is the immediate and undeniable danger of terrible suffering. As far as patients are concerned, the majority of cases will probably be different, since the reason why they want to die might not be as immediate or obvious. Despite this, the danger of

terrible suffering or loss of dignity, which are neither as immediate nor evident, is not any less real. If we accept this, we should see no significant difference between them.

Therefore, we should focus only on the differences that bear a moral relevance. Thus, those criticising the second premise should identify a morally relevant feature, F, which is present in one case but missing from the other. In order to provide support for the second premise, I consider two possible candidates for F; *terminal illness or mortal danger* and *physical suffering*. (To get an overview of classic legal conditions for candidacy for euthanasia, see Young 2019.)⁴

In the case of the blazing lorry, the driver found himself in mortal danger and the likelihood of him having to endure great physical suffering was very high. Should we thus presume that physician-assisted euthanasia is only permissible in cases when the patient's life is in danger (or when they suffer from a terminal illness) and only when their suffering is of a physical nature? I do not think that the morality of euthanasia should depend on whether there is terminal illness, mortal danger, or physical suffering involved.

Where terminal illness and mortal danger are concerned, there are patients whose incurable diseases put them through intense agony, and they wish to die even though their illness is not life-threatening. These patients suffer greatly and can reasonably conclude that their lives are not worth living. Why should we consider the fact that they are not terminally ill to be morally relevant?

One of the reasons why the insistence on conditions such as terminal illness or mortal danger might seem necessary is because there is a possibility that the patient's condition might improve in the future—for instance, with the invention of a new revolutionary cure or as a consequence of unexpected remission. I believe that in particular cases, these factors should be taken into account when evaluating the rationality of euthanasia. However, they should not be regarded as an obstacle to performing euthanasia on non-terminally ill patients or people not finding themselves in mortal danger. The reason behind this is that if a patient is neither terminally ill

⁴ Naturally, someone might always point out a morally relevant feature that differentiates requested euthanasia from the trapped lorry driver's case. If an important difference has escaped my attention, I would like to be informed about it.

nor in mortal danger, but their life is still not worth living, more suffering could be avoided with euthanasia, since without it they are likely to live longer and suffer even more. If we imagined, *per impossibile*, that the lorry driver would not be burning alive for several minutes but for days, months, or even years, we would have all the more reason to spare him this suffering. Nevertheless, if we or our patient had a good reason to believe that there is a realistic chance of improving his or her situation, I do not think we should comply with his or her request for euthanasia. However, we can come to understand with a reasonable amount of certainty that some patients do not have a realistic chance.

The next condition relates to intolerable physical pain, which is put in direct contrast with mental or psychological pain. As far as I can see, we should not differentiate between these types of suffering where euthanasia is concerned, as psychological suffering is not any less real or unpleasant than physical one. Despite this, I maintain that when patients suffering from psychological pain rather than physical pain request euthanasia, we should deliberate their petition with much more care.

I suppose that the insistence on the condition of physical pain is similar to the previous one. It is more difficult to assess whether a patient's life has ceased or will cease to be worth living if they suffer from psychological pain than if they experienced physical pain; the former is much more elusive. I find this to be a good reason for regulating euthanasia more strictly when it comes to patients afflicted with psychological pain; however, this does not mean that euthanasia should be denied to them. Some people suffer from incurable and unbearable psychological pain and their psychological condition renders their lives not worth living. If we cannot help them in any way and we have no reason to believe that we will be able to help them in the future, we might consider complying with their request.

I do not mean to suggest that patients should be given euthanasia without trying to help them first. On the contrary; everyone requesting euthanasia should be offered help, whether in the form of psychiatric therapy, palliative care, or some kind of experimental treatment. What I am suggesting is that those not in mortal danger or whose suffering is mainly of a psychological nature should be presented with stricter legal conditions when requesting euthanasia.

Although I consider these arguments convincing, others might not. It is also likely that some of my claims might prove to be untrue or unacceptable. If I am mistaken, I would like to be informed on the fallacies in my thinking, which I presume there will be; even in our provenience many objections have been raised against the position I am trying to defend. In the next part of this study, I will try to explain why my opinion on euthanasia has not been changed by them.

* * *

Objection 1: *The right to die is beyond the limits of the freedom of an individual. Death is something humans have no moral right to decide about. Since euthanasia involves one person killing another, it is not permissible.*

Response: Saying that one's right to die violates the (moral) limits of personal freedom does not seem right, since our society does not object when someone voluntarily puts their life in danger. We do not denounce people who risk their lives by becoming soldiers, police officers, fire-fighters, or stunt performers, nor do we condemn those putting their health in jeopardy with their lifestyle choices—say, by damaging their lungs by smoking or by pursuing extreme sports. This means that while a person is free to put their life at risk and lose it in the case of an accident, or consciously shorten and endanger it with the lifestyle they lead, they are not free to end it directly. Would those who raise this objection morally condemn all activities whose pursuit puts a person's life at risk or shortens it?

Our society accepts such behaviour, and there are even instances when a person voluntarily taking their life is thought to be highly commendable. For example, a soldier throwing himself on a grenade and saving his friends is not committing a morally deplorable act but a laudable one! The perception of a mother dying for her child is similar.

An opponent of euthanasia might suggest that when someone requests the termination of their life, they do not sacrifice themselves to save another person's life—they do so for their own well-being. However, I do not find this objection convincing enough. We could assume that the soldier's sacrifice did not save his friends' lives but "merely" spared them years of terrible torture. We could similarly imagine that the mother did not sacrifice herself to save her child's life but to stop it from being brutally tormented. I suppose

most of us would not accuse these people of doing anything wrong. If we were to morally evaluate their sacrifice, we would be more likely to consider it laudable.

Jakub Jirsa (2011, 587) wrote in his paper that “my life is something I do not have right to do with as I see fit,”⁵ hence ending a life prematurely is morally problematic. I assume Jirsa is talking about a moral right rather than a legal one. I will furthermore assume that breaking a moral right means doing something wrong. If that is so, then Jirsa’s sentence suggests that *ending or trying to end one’s life is not permissible*. If he were right, it would mean that neither suicide and physician-assisted suicide nor requested euthanasia were permissible, since they involve either ending one’s life or an attempt to do so.

In my interpretation, Jirsa offers two different arguments to support his claim about the limits of a person’s moral freedom. Let us start by considering the first one, which is based on the assumption that “I am permitted to do what I see fit only with what is (or could theoretically be) entirely within my control” (Jirsa 2011, 588). The author complements this contention by saying that one’s own life is not something one entirely controls. Thus:

Jirsa’s first argument

P1 My life is not and theoretically could not be entirely within my control.

P2 I am morally permitted to do as I see fit only with what is or theoretically could be within my full control.

C Therefore, I am not morally permitted to do with my own life as I see fit.

Let us look at the premises. What justifies the first premise? To support it, Jirsa is citing Galen Strawson, who argues that *we cannot prove to be truly morally responsible for our actions*. For the sake of concision, I rephrased Strawson’s argument as follows:

⁵ All translations from Czech are my own.

Strawson's argument

- P1 "You do what you do because of the way you are."
 P2 "To be truly morally responsible for what you do you must be truly responsible for the way you are—at least in certain crucial mental respects."
 P3 "You cannot be truly responsible for the way you are, so you cannot be truly responsible for what you do."
 P4 "To be truly responsible for the way you are, you must have intentionally brought it about that you are the way you are, and this is impossible." (Strawson 1994, 13–14)
 P5 It is impossible to intentionally bring about who you are, because you must have existed before beginning to exist.

C Thus, you are not morally responsible for what you do.

Strawson's argument denies the existence of moral responsibility. If we accepted his contention, we would have to believe that it is not possible to be morally responsible for anything. Strawson himself clearly does not take this conclusion seriously; he presents it as a philosophical puzzle. His aim is to provoke a defensive reaction in philosophers so that they present a satisfactory explanation for the nature of moral responsibility without resorting to such absurdities, as there is no moral responsibility.

Jirsa's first premise most likely seeks support in premises of Strawson's argument. The way I am defines how I behave and decide. But since I have not intentionally brought about the way I am—as it is impossible—I cannot have any control over the way I am, and as a result I cannot have any control over what is defined by my nature—i.e., my behaviour and decision-making.

However, if this is the reasoning behind Jirsa's first premise, it means that we have no control over anything. No decision or act would be free. For the sake of argument, let us assume that Jirsa's first premise is true.

Let us take a look at the justification the second premise depends upon. Jirsa claims (Jirsa 2011, 588) it to be his assumption. For the sake of argument, let us now presume it true. If we recognised this statement, we would not be morally permitted to do as we see fit with anything. Because by accepting the justification Jirsa provides for his first premise, we would also

have to acknowledge that we do not have full control over anything. But, as the second premise suggests, if having full control over X is the precondition for doing with X as we see fit, then we do not have the moral permission to do as we see fit with anything.

I would like to contradict this by saying that I am morally permitted to do as I see fit with my choice of words in this sentence. I find the implication that moral permission is impossible to be unacceptable. We should rather refuse Jirsa's line of reasoning than accept such an implication. Nevertheless, if we carried on analysing Strawson's puzzle, we would stray too far from our discussion on euthanasia.

Jirsa's second argument is based on Wittgenstein's thesis from his *Tractatus Logico-Philosophicus*, declaring that "[d]eath is not an event in life but is the end of life" (qtd. in Jirsa 2011, 588). Jirsa's conclusion was that *ending or trying to end one's life is not permissible*. However, the claim that death is not an event in life but is the end of life does not imply that ending or trying to end one's life is morally wrong. We thus need to complement it with an additional premise.

I am not sure which one Jirsa would prefer, since he did not express it explicitly. I will assume that since Jirsa believes that death is not an event in life, and that therefore it is wrong to bring it about, than *it is wrong to bring about everything that is not an event in life*. Based on this premise, we can construct the following argument:

Jirsa's second argument

P1 If X is not an event in life, then it is wrong to bring about X.

P2 Death is not an event in life, it is the end of life.

C Therefore, it is wrong to bring about death.

If this is Jirsa's second argument, then I do not find it convincing. If a person cannot decide on the death of another person because death is not an event *in the life* of the person to be killed but is the end of it, then, for the same reason, they cannot decide about the beginning of life. Conception is also not an event in the life of the conceived—it is the beginning of it. Physicians would thus not be permitted to provide assistance not only in death, but in birth as well.

My last objection against Jirsa's position is independent of the argument provided for its support. If the position was true or acceptable, then shooting the trapped lorry driver, together with the soldier's and mother's sacrifice, would be wrong. Should we accept this conclusion? I do not think so. This is why I believe Jirsa's arguments should not be accepted. However, it is possible that I misinterpreted or misunderstood his reasoning.

Objection 2: *Euthanasia should be prohibited as it would lead to a slippery slope. If we started practising it, it would eventually result in involuntary euthanasia or even medical murder.*

Response: The slippery slope objection fundamentally differs from the previously proposed arguments. Although it claims that we should not be practising euthanasia, it does not say that euthanasia itself is wrong. What it deems morally wrong are the side effects of practising and legalising euthanasia. Euthanasia is thus *indirectly* wrong.

Those against the legalisation of euthanasia maintain that we would find ourselves on a slippery slope that would eventually lead to involuntary euthanasia and medical murder. These critics worry that by legalising requested euthanasia, sooner or later we would due to sociological, psychological, or even logical reasons end up emulating practices of Nazi Germany. They worry that if euthanasia became an option, it would eventually turn into an expectation or even a requirement.

I presume that we would not like to live in a society that required or even forced some of its members to undergo premature and unwanted death; I know I would not. How then can I advocate for the legalisation of requested euthanasia?⁶ Marta Munzarová (2012, 416) described the situation in the Netherlands in the following way: "Killing patients without their request is still happening, but the reporting must be different (since it is not euthanasia, which is defined by Dutch law as killing 'on the patient's request'). Can we imagine a more illustrative example of a slippery slope?"

I support the legalisation of euthanasia because I do not believe it would lead us down a slippery slope. Firstly, advocating for requested euthanasia

⁶ I support the right to euthanasia in the sense that a patient should have the option to receive euthanasia, but I do not support it in the sense that a physician should be obliged to administer it to them.

and rejecting involuntary euthanasia is not contradictory. The distinction between them is clear. The cases Munzarová mentions show how patients' lives were terminated without their explicit request. The authors of the study claim that these were predominantly patients who were unable to express their will because, for instance, they were in a coma. These were instances of so-called non-voluntary euthanasia,⁷ usually performed on patients who were incapable of giving voice to their decisions; thus, these patients did not die against their will.

I suppose Munzarová is trying to point out that the medical records relating to the termination of lives contain a category of patients who did not explicitly consent to being killed. This fact in itself does not prove that there is a slippery slope; only long-term statistics could confirm that, as one-year data do not adequately illustrate how many cases of euthanasia without the explicit consent of the patient were performed in previous years. We should compare records of several years and see if they indicate a growth in deaths we consider wrong. Moreover, we also do not know how many instances recorded in the category of "a patient's life ended without their explicit consent" occurred in countries that still see euthanasia as murder but find terminal sedation acceptable (like Slovakia, for example), since we do not keep a record of them. As a result, we do not know if a system that has legalised euthanasia is better or worse in this aspect than a system where it is illegal.

But let us get back to the key question. Is there any evidence to support the contention that in the Netherlands, instances of unacceptable—or less acceptable—forms of ending lives are on the rise? A Dutch study conducted in 2010 (Onwuteaka-Philipsen et al. 2012, 912) indicated a decline in the frequency of cases where lives were terminated without the explicit consent of the patient, while a meta-analysis of Dutch studies (Rietjens et al. 2009, 279) concluded that "the legalization of euthanasia in the Netherlands did not result in a slippery slope for medical end-of-life practices." Therefore,

⁷ There is a difference between involuntary and non-voluntary euthanasia. In the case of involuntary euthanasia, the patient does not want to die, or they can express their will but it is being ignored. In the case of non-voluntary euthanasia, the patient is unable to communicate what they want.

there is no evidence to support the claim that the Dutch euthanasia practices are on a slippery slope.

In addition, I would not like to brush off Munzarová's suggestion that non-voluntary euthanasia is inherently morally wrong—if that is indeed what she suggests. I believe that there are situations—especially those involving people who will never regain consciousness—where non-voluntary euthanasia should be permitted. However, I will have to defend that thesis on another occasion.

Finally, if we eventually found out that in spite of what current records indicate, the legalisation of euthanasia would lead our society to a slippery slope, we can always change the law and reinstate the former system. So even if legalising euthanasia proved to be a mistake, we would not be obliged to continue with it forever. Therefore, I do not think there is a reason for us to worry about finding ourselves on a slippery slope.

There is, however, one additional argument, somewhat close to the slippery slope idea. One could argue that legalising euthanasia would be harmful even if no one was forced to undergo it, because it would necessarily present the population with an uncomfortable choice. The mere fact that one is presented with a choice between continued life and euthanasia could exert, in some segments of the population, a measure of psychological pressure and therefore cause suffering. I believe this reasoning is close to Jirsa's line of argument, since he has voiced his opinion that legalising euthanasia would exert pressure on patients that are the most vulnerable and disadvantaged (Jirsa 2011, 589).

In response, I would like to say, somewhat vaguely, that it is far from clear that the value gained by avoiding suffering caused by the pressure of choice is greater than the value gained by avoiding suffering of euthanasia applicants. I personally doubt it is. Although I have no evidence for this claim, I presume that the proponents of this argument would agree with me that this argument, just like the slippery slope argument, should be based on empirical evidence.

Objection 3: *Euthanasia should not be practised as there is no need for it. What renders it needless is the efficiency of palliative and hospice care, along with the fact that patients may refuse nutrition and hydration and will thus die without any assistance.*

Response: Some claim that there is no need for euthanasia because palliative and hospice care have improved so much that they can significantly reduce the patients' suffering. Marek Vácha (2010, 273) maintains that "palliative care has advanced so much that when a patient has been suffering from unbearable physical pain for a long time, he is most likely receiving the wrong treatment." Jirsa seconds this opinion, as does Munzarová (Jirsa 2011, 581; Munzarová 2012, 416).

This objection differs from the previous two as it refrains from calling euthanasia directly or indirectly wrong and merely attempts to prove its needlessness;⁸ therefore, it does not challenge my thesis about requested euthanasia being morally permissible. Despite this, I would still like to address this objection, as it has been presented by several authors.

I do not mean to imply that there is anything wrong about palliative or hospice care. I do believe that it is important for people whose quality of life is expected to decline to avail themselves of such avenues of treatment if they want to. We should fully support those who make use of such treatments rather than wishing to die, as every potentially helpful option should be examined and tried. This is why I would prefer it if palliative care received wider recognition.

However, it is up to the patient to ultimately decide if they find palliative care useful; they can judge for themselves whether it makes their life worth living or not. After all, it might not; contrary to Vácha's view, nothing guarantees that palliative care will render their lives more liveable (unless my opponents can prove that it is 100% effective). Even if palliative care could relieve a patient of their pain, it might come at a higher cost than the patient is willing to pay. By the time the physician successfully identifies a pain management method that would suit a particular patient's needs, it is possible that the patient will have had to suffer through a lengthy period of trial and error during which their pain will not be alleviated and their quality of life will not improve. And even if the physician found a suitable treatment, the patient might spend the rest of their life experiencing nausea, incontinence, frequent losses of consciousness or other

⁸ This objection is more pertinent to debates disputing the rationality of euthanasia rather than its morality; however, I decided to include it, as it is normally presented as a moral reproach.

distressing symptoms. If a patient decides that they do not want to receive palliative care at such a cost and refuses to spend their remaining days with such poor quality of life, they might want to die with medical assistance. This is supported by studies of patients who have been presented with a choice between high standard palliative care and a physician-hastened death and who preferred or would prefer the latter (Wilson et al. 2007; Quill et al. 2008).

This objection differs from the others in one further aspect. Rather than attacking the permissibility of euthanasia, it challenges its legalisation. I do not mean to claim that if something is morally permissible then it must also be legal. However, I do think that if an act is morally permitted then the most fundamental objection against its legalisation loses its ground. Those opposing the legalisation of euthanasia carry the burden of proof in showing that as a society, we should legally prohibit something that is morally permitted. I believe that Jirsa's remaining objections could be solved if euthanasia was properly regulated, as no supporter of sound mind would want to legalise euthanasia without taking proper regulatory measures.

Legalisation and morality are two different things and I fully agree with Jirsa (2011, 586) that we should not lose sight of what sets them apart. I think that in some countries, legalising euthanasia might cause more harm than good. Similarly to Jirsa and several other participants of the euthanasia discussion in Slovakia, I worry that legalising euthanasia under the circumstances that currently prevail in the country would be rather harmful, although I do not know it with certainty. I do not know if legalising euthanasia would ultimately have a negative impact on our society, as scientific evidence is needed to get a well-founded answer to this question.⁹ To the best of my knowledge, there is no evidence for any of these sides, which is why I refuse to take a stand in this discussion. But even if we accepted that

⁹ If I may venture a speculation, the experience so far suggests that physician-assisted death and euthanasia are generally being implemented by developed countries with high healthcare standards (Physician-assisted death: Switzerland, Oregon, Washington, Montana, Vermont, Canada, and soon Australia; Euthanasia: The Netherlands, Belgium, Luxembourg, and Colombia). This might signify that implementing such end-of-life practices in countries with less developed health care systems could lead to adverse consequences.

in the current state of affairs euthanasia would be harmful, conditions might improve in the future, which might present the right circumstances for legalising euthanasia. Besides, if we believe that euthanasia is morally permitted and should be legally accessible, we should take active means to bring about such a future.

* * *

Summa summarum, I think that from a moral point of view, the requested euthanasia of adults is permissible, because I accept the argument from analogy, and because I am not aware of any good objection against its permissibility. Where the legal aspect is concerned, I differentiate between immediate legalisation and legalisation when the time is right. Since I do not know if the current conditions in Slovakia or in other states are appropriate for legalisation, I take the position that euthanasia should be legalised once the conditions are ripe for it. In a nutshell, euthanasia should be legal—if not at this moment, then sometime in the future. But even if euthanasia should be legal sometime in the future, we should take proactive steps to create suitable conditions for its implementation.

Funding

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-14-0510.

References

- Dworkin, Ronald. 1993. *Life's Dominion: An Argument about Abortion, Euthanasia, and Individual Freedom*. New York: Alfred A. Knopf.
- Greif, Adam. 2018. "Racionálnosť eutanázie" ["Rationality of Euthanasia"]. *Časopis zdravotníckého práva a bioetiky* 8 (2): 43–58. <http://medlawjournal.ilaw.cas.cz/index.php/medlawjournal/article/view/168/142>
- Hare, Ronald M. 1975. "Euthanasia: A Christian View." *Philosophic Exchange* 6 (1): 43–52. https://digitalcommons.brockport.edu/phil_ex/vol6/iss1/2/
- Hope, Tony, Savulescu, Julian, and Hendrick, Judith. 2008. *Medical Ethics and Law: The Core Curriculum*, 2nd Edition. Edinburgh: Churchill Livingstone.
- Hříbek, Tomáš. 2010. "Za etiku bez teologie" ["For an Ethics without Theology"]. *Filosofický časopis* 58 (5): 729–49.

- Jirsa, Jakub. 2011. "Problémy s asistovanou sebevraždou" ["Problems with Assisted Suicide"]. *Filosofický časopis* 59 (4): 579–90.
- Wilson, Keith G., Chochinov, Harvey Max, McPherson, Christine J., Skirko, Merika Graham, Allard, Pierre, Chary, Srini, Gagnon, Pierre R., Macmillan, Karen, De Luca, Marina, O'Shea, Fiona, Kuhl, David, Fainsinger, Robin L., Karam, Andrea M., and Clinch, Jennifer J. 2007. "Desire for Euthanasia or Physician-Assisted Suicide in Palliative Cancer Care." *Health Psychology* 26 (3): 314–23. <https://doi.org/10.1037/0278-6133.26.3.314>
- Luper, Stephen. 2009. *The Philosophy of Death*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9780511627231>
- Munzarová, Marta. 2012. "Proč NE eutanazii" ["Why We Should Say NO to Euthanasia"]. *Filosofický časopis* 60 (3): 403–20.
- Onwuteaka-Philipsen, Bregje D., Brinkman-Stoppelenburg, Arianne, Penning, Corrine, de Jong-Krul, Gwen J.F., van Delden, Johannes J.M., and van der Heide, Agnes. 2012. "Trends in End-of-Life Practices Before and After the Enactment of the Euthanasia Law in the Netherlands from 1990–2010: A Repeated Cross-sectional Survey." *The Lancet* 380 (9845): 908–15. [https://doi.org/10.1016/S0140-6736\(12\)61034-4](https://doi.org/10.1016/S0140-6736(12)61034-4)
- Quill, Timothy E., Lo, Bernard, and Brock, Dan W. 2008. "Palliative Options of Last Resort: A Comparison of Voluntary Stopping Eating and Drinking, Terminal Sedation, Physician-Assisted Suicide and Voluntary Euthanasia." In *Giving Death a Helping Hand*, edited by Dieter Birnbacher and Edgar Dahl, 49–64. Dordrecht: Springer. <https://doi.org/10.1007/978-1-4020-6496-8>
- Rietjens, Judith A.C., van der Maas, Paul J., Onwuteaka-Philipsen, Bregje D., van Delden, Johannes J.M., and van der Heide, Agnes. 2009. "Two Decades of Research on Euthanasia from The Netherlands: What Have We Learnt and What Questions Remain?" *Journal of Bioethical Inquiry* 6 (3): 271–83. <https://doi.org/10.1007/s11673-009-9172-3>
- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75 (1–2): 5–24. <https://doi.org/10.1007/BF00989879>
- Vácha, Marek. 2010. "Je vůbec ještě možná etika v 'postetickém' světě?" ["Is it at all Possible to Have Ethics in a 'Post-ethical' World?"] *Filosofický časopis* 58 (2): 273–79.
- Young, Robert. "Voluntary Euthanasia." *The Stanford Encyclopedia of Philosophy* (Spring 2019 Edition), edited by Edward N. Zalta. Last updated January 18, 2019. <https://plato.stanford.edu/archives/fall2017/entries/euthanasia-voluntary/>


Historical Antirealism and the Past as a Fictional Model

David Černín*

Received: 15 June 2018 / Accepted: 28 January 2019

Abstract: This paper focuses on the discipline of history, its methods, subject, and output. A brief overview of contemporary analytic philosophy of history is provided, followed by critical discussion of historical realism. It is argued that the insistence on the idea that historians inquire into the real past and that they refer to the actual past entities, events, or agents is widely open to sceptical objections. The concept of an abstract historical chronicle of past events which are explained or retold by historians is identified as misleading. The idea of historical antirealism is then introduced. It is argued that in the centre of historian's attention are present phenomena that are identified as historical evidence and require historical explanation. Historical explanation consists of constituting an historical past—a fictional model that accounts for present data. The identification process of historical evidence and the discursive nature of historical enterprise are analysed and accompanied by several concrete examples. According to historical antirealism, historians are not interested in the real past, but in the present empirical data. In their pursuit of historical knowledge, they produce fictional models—an historical past. Lastly, several common caveats against historical antirealism are addressed. The historical antirealism is presented as a viable fictionalist account of the historical inquiry that is capable of avoiding

* University of Ostrava

 FF OU – The Centre for Research in Medieval Society and Culture (VIVARIUM);
Department of Philosophy, Faculty of Arts, University of Ostrava, Reální 5, 701
03 Ostrava, Czech Republic

 cernin.d@gmail.com



sceptical attacks on historical method and it is argued that antirealism allows history to retain its worth as a distinctive kind of scientific discipline.

Keywords: Analytic philosophy of history; antirealism; evidence; explanation; Goldstein; history; model.

1. Outset

This paper aims to reassess the philosophical conceptions of historical inquiry, to identify some problematic ideas that are more or less present in contemporary philosophy of history (historical realism), to explore controversial alternative (historical antirealism), its relations to fictionalism, and to highlight its possible merits. The paper proceeds by the introduction of the analytical philosophy of history followed by the assessment of historical realism and historical narrativism, which introduces fictional elements to the discourse. Then it moves to the preferred alternative—the historical antirealism and the final part answers the most common caveats against the historical antirealism.

A fictionalist approach to the wide range of entities involved in the process of scientific explanation is a notoriously attractive standpoint. Certain entities (e.g. highly idealised models of atoms, clusters of space bodies, social institutions, or historical periods) may be regarded as fictional by philosophers. Nonetheless, as long as these models¹ allow scientists to achieve an understanding of our world and present phenomena, to test their assumptions or predictions by the means of controlled experiments, to explain various observations which are accommodated by the model, or to produce results affecting our daily practices, we generally consider these fictions to be useful and therefore vindicated. Fictional and idealised models are often in the very centre of scientific discussions and they are continuously changed or adjusted to better account for the phenomena in question. This fluid nature of scientific models compels some philosophers to treat

¹ Models in this paper are understood as ‘a work of fiction’ in the same sense as presented by Nancy Cartwright, see (Cartwright 1983, 153). For further discussions, see (Hartmann et al. eds. 2008).

them as fictional entities, relieved of the ontological burden, which play important roles in many scientific disciplines.

Fictional entities help scientists to apprehend and simplify complex systems, to make predictions, to mediate the fruits of their research process to the public, or they can be employed and utilised by other scientific disciplines. The very act of constituting a fictional entity is scarcely the final step for researchers to do, since creating a fiction is certainly not regarded as the goal of any scientific endeavour. However, some disciplines can become easily contested in this respect. A fictionalism and its associated issues are often brought up in relation to various disciplines that are focused on the past entities and events. This entails not only the history² and historiography but also some natural sciences and subfields, like cosmogony, geology, evolutionary biology, etc.³ The former group is studied by philosophers of history and philosophers of historiography,⁴ while the latter is the subject of the philosophy of science. The parallels between natural sciences and human history are continuously discussed by scientists and philosophers.⁵ At the same time, philosophers of historiography appreciate

² By history in this paper I mean the discipline of history and historical inquiry, not the course of past events itself. Cf. (Tucker ed. 2009, 1–6). In this context, it is therefore meaningless to ask questions about the meaning of history, the destination of history, or about its predictive capabilities. The real course of past events will be referred to simply as the past.

³ This idea is clearly expressed by David L. Hull for whom the cosmogony, geology, palaeontology, and human history ‘are the four most important historical disciplines’ (Hull 1975, 264).

⁴ By philosophy of history or philosophy of historiography, I mean especially contemporary Anglo-Saxon philosophical tradition which aims at explaining the practice of history. This discipline is variously called analytic philosophy of history, epistemological philosophy of history, critical philosophy of history, or philosophy of historiography. It is, therefore, very different from substantial or speculative philosophy of history, which entails authors like Francis Fukuyama, Arnold J. Toynbee, or Karl Marx. This paper does not focus on the speculative philosophy of history.

⁵ The philosopher of biology David Hull and his 1975 article is a great example. He also joined discussions about the aims and methods of the history of science in (Hull 1979). Prominent evolutionary biologist Richard Lewontin also pointed out

the insight into scientific practices brought by prominent philosophers of science like N. Cartwright or W.V.O. Quine and they try to apply their conclusions to the history and historiography.⁶

The historical enterprise, its goals, methods, rules, merits, and limits—all those lie in the centre of many philosophical disagreements. At first glance, it may seem trivial to claim that the history focuses on the past and the historians are supposed to discover and explain the past facts to their audience. Historians are interested in the actions of historical agents and they should mediate their reasoning to readers. They are supposedly meant to find the reasons for wars, revolutions, migrations, unexpected victories in elections, or economic crises. This entails the field of historical explanation—the longstanding debate, which can trace its origins to the C.G. Hempel’s influential paper “The Function of General Laws in History” (Hempel 1942). Nonetheless, it has been contested that works of historians do not represent the most illustrious cases of covering law application.⁷ Still, the interest in the components of historical explanation lasts to these days, it often intersects the narrativist philosophy of history,⁸ and it is also echoed

striking similarities between human history and evolutionary biology in (Lewontin 1991). See also (Wilkins 2009).

⁶ Aviezer Tucker, an influential philosopher of historiography, draws a lot of inspiration from the philosophy of biology and he even joined their discussions; see paper (Tucker 2011). He also approaches the study of historiography with the concept of Quine’s naturalized epistemology in mind; see (Tucker 2001, 49) or (Tucker 2004, 9). M.G. Murphey mentions N. Cartwright as the important influence on his work and the reason for his abandonment of Hempelian paradigm; see (Murphey 2009, x). The analytic philosopher Paul A. Roth also focuses both on natural sciences, Quine’s intellectual legacy, and the narrativist philosophy of history at the same time.

⁷ (Hempel 1942) was written by Hempel shortly after his emigration to the USA. This paper has drawn the attention of many Anglo-Saxon philosophers to the discipline of history and it is discussed to this day. See (Mandelbaum 1961; Wright 1971; Murphey 1986), etc. Contemporary historical accounts of the analytic philosophy of history usually start with this paper and critical reaction to it, see (Ankersmit 1986; O’Sullivan 2006; Kuukkanen 2015).

⁸ By narrativist philosophy, I refer to the philosophical tradition represented by W.B. Gallie, A.C. Danto, H. White, F. Ankersmit, etc.

in the discussions concerning historical explanation and the role of general laws in natural sciences.⁹

The narrativist philosophers of history often herald their own linguistic turn in the philosophy of history and they openly focus on the language historians use.¹⁰ The act of historical writing and the linguistic structure historians impose on the past stand in the centre of their attention. Narrativists stress the choices historians make when presenting their accounts of historical events. Historical enterprise is basically a kind of literature and storytelling while a spatiotemporal and cultural background of the historian—storyteller vastly influences the final outcome. Narrativism thus allows for a diversity of accounts we find in the field of history and historiography and which we cannot easily disregard.

Nonetheless, even narrativist philosophers of history do require some connection to the past historians are writing about. In order to provide a narrative explanation, historians and scientists¹¹ must possess some knowledge of the past events and facts that are chronologically ordered into sets, possibly in the form of singular factual statements. These basic sets are a necessary prerequisite for any narrative explanation and they are generally called “a chronicle” by narrativist philosophers.¹² Term “chronicle,” meaning singular non-causal factual statements as a prerequisite for writing explanatory history or narrative, is also to be found in the works of Morton White (see White M. 2005, 40).

⁹ E.g. Alexander Rosenberg stresses the fact that historical explanation is a specific case of explanation in (Rosenberg 2007, 129–31); see also (Rosenberg 2000). Rosenberg is known for being strongly critical of “philosophers of history” and “history buffs,” however, his criticism usually pertains to the speculative philosophy of history and he does not refrain from borrowing concepts and distinctions from critical philosophy of history which is “a division of philosophy whose relevance to biology may now be apparent” (Rosenberg 2000, 151).

¹⁰ An exemplary historical narrative of the narrativism and its development is provided by (Ankersmit 1986).

¹¹ Paul Roth argues that even evolutionary biology employs narrative explanations since every retrospective explanation (e.g. how an adaptive mutation came to be) is an essentially narrative explanation which must necessarily proceed by the chronological sequencing of preceding events, i.e. creating a narrative. See (Roth 2017).

¹² See (White H. 1973). For critical discussion, see (Murphey 2009, 103–34).

The narrativist philosophers of history use the distinction between a *chronicle* and a *story* or *history* to highlight the main focus of their approach to historical enterprise. The mere chronological ordering of events does not constitute a proper historical account, it does not explain anything, and does not tell a story:

The story transforms the events from the meaninglessness of their serial arrangement in a chronicle into a hypotactically arranged structure of occurrences about which meaningful questions (what, where, when, how, and why) can be asked. (White H. 1975, 59)

Even Hayden White compares this task of historians to the creation of model:

But this fashioning is a distortion of the whole factual field of which the discourse purports to be a representation—as is the case in all model-building. (White H. 1975, 60)

This layer of model-building (the transformation of a chronicle into proper history) lies in the centre of narrativist's attention and it does pose a truly fascinating theoretical issue on its own.

However, the very idea of a chronicle of past events, providing building blocks for historical narratives, should be particularly troubling for historians and scientists. Neither evolutionary biologists nor historians will ever have the full “chronicle” of past entities at their disposal. This chronicle itself could be considered as an idealisation or fiction and as such, it also begs an important question: How do we obtain the knowledge of the past events and entities? Is it a knowledge of the real past? Is a chronicle factual or already fictional? Consequently, the very same question pertains also to the discussion about historical explanation. It will be shown that this is not merely a question regarding the context of discovery of sole historical fact. On the contrary: it is a question about the more fundamental level of model-building that precedes the formation of a chronicle.

2. Historical realism and invitation for scepticism

Both narrativist philosophy of history and analytic philosophy of history still show some residual signs of naïve historical realism when accounting

for the discipline of history. This can be quite surprising in the case of narrativist philosophy, which does offer a more relativistic portrait of historians exercising their liberties when writing historical narratives. In the words of Frank Ankersmit: “the historical narrative is a complex linguistic structure specially built for the purpose of showing part of the past” (Ankersmit 1986, 19). At the same time, Ankersmit rejects naïve historical realism as inherently flawed as well as the majority of contemporary philosophers and theoreticians of history.¹³

Historical realism does invoke conception akin to the correspondence theory of truth. Historians are considered to be inquiring into the real past and the touchstone for their theories are real past events as they actually happened. This may even appear as an unproblematic statement: historians inform us about events, people, cities, buildings, pieces of art, organisms, and generally about the entities we encounter in everyday life and we consider them to be real. None of these objects is abstract in any way and we do not require historians to describe Caesar as a real human with two legs and two arms so we can imagine him properly. The majority of historical events is observable in principle (see Murphey 2009, 10–14) and therefore we can imagine them as represented by historians and historical texts. According to the historical realists, historical explanation, in any of its form, is bound to explain the processes in the past.

On the other hand, nobody would ever claim that historians do have any mysterious access to the past.¹⁴ Historians do not time-travel and they

¹³ There are various philosophical approaches to a history that try to avoid naïve historical realism: Some philosophers of history do not consider the discussions about historical realism/antirealism relevant, see (Tucker 2001, 52), while others subscribe to moderate versions, like “constructivist realism,” see (Murphey 2009, 13). Maurice Mandelbaum could be portrayed as defending position akin to historical realism in the case of “general history”, but not in the case of “special histories” (Mandelbaum 1977). Although historical realism is scarcely defended, there are some exceptions—e.g. Sir Geoffrey Elton, an historian whose texts on the historical method are considered to favour historical realism.

¹⁴ There is, of course, a rather problematic case of R.G. Collingwood and his notion of “re-enactment,” which is commonly interpreted as a metaphysical link to the thought of historical agents, through which historians can re-think or re-live the past, see e.g. (Tucker 2004, 200–207). However, there are also interpreters who

do not observe the events they are writing about. At the same time, they do not (generally speaking) conduct experiments. There are, of course, exceptions to both of these statements: historians can write about the events they themselves experienced (observed),¹⁵ they can work (in the case of recent history) with movie footage, or they can interview living witnesses.¹⁶ They can even employ the methods of experimental archaeology which make use of laboratory equipment and modern methods to test and replicate various artefacts our ancestors were apparently using during their lifetime. Various types of chemical and spectrographic analyses of historical relicts are also widely used to recover significant data. All of these methods are genuinely historical, they are successfully and critically employed, and neither of these invites any fatal criticism which could uncover it as profoundly unscientific. However, none of these methods does interact directly with the real past in any way.

At the same time, many entities that are postulated by historians and that are said to affect past events are admittedly abstract constructs which defy precise definition. Ancient Greek philosophy is a common term in many texts inquiring into the history of philosophy. Nonetheless, it does not aspire to describe all Ancient Greeks who have ever philosophised. Neither does it seek to explain all present artefacts (especially texts) which are identified as instances of this philosophy. Historians of philosophy are rather trying to refine the contemporary notion itself upon the basis of a limited number of selected texts and the very process of selection of canonical texts is an argumentative strategy in itself (see Guérout 1969). Unanimous consensus on the canonical texts and their interpretations does not exist but we can still talk about the Ancient Greek philosophy as influencing future texts,

understand Collingwood more in lines with constructivism and even antirealism; see (Nielsen 1981), (Dussen 2012), or (Goldstein 1996).

¹⁵ However, simply “remembering something” is definitely not the same as “having an historical account of something.” Historical inquiry entails critical approach to evidence and memories of witnesses or historians can be used as such evidence.

¹⁶ The oral history has come a long way since the time of Herodotus. Interviewing witnesses and contemporaries is now a highly systematic process which takes a heed of other disciplines, including psychology, in order to filter out possible personal biases, etc.

authors, or ideas. Many terms commonly used to denote specific periods like the Renaissance, the Cold War, the Thaw, or the Thirty Years' War have been analysed by the philosophers of history. These colligatory terms help us to make the past more comprehensible, even though their clear delimitation is not firmly given.¹⁷

Some philosophers have tried to answer the caveats against historical realism by compromise. Maurice Mandelbaum established a distinction between general history and special histories (Mandelbaum 1977, 162–65). He holds that general history is interested in the continual existence of a society (e.g. France), while special histories (e.g. French literature of certain period) are interested in the collection of separate works, whose identification is theory-laden. These considerations lead him to a conclusion that special histories cannot claim objectivity in any way. However, the claim that historians recognise societies as entities having continuous existence has been contested by other philosophers of history.¹⁸ Michael Oakeshott directly states that it is easy for philosophers to challenge even the seemingly unambiguous notions like England since the subject of history is not a *datum* but it is established in the course of historical inquiry and the identification of its subject is entirely in the hands of the historian (Oakeshott 2004, 404–406). Contemporary historical accounts of any European country are narratives of many different societies that were radically changing as well as their own understanding of historical identity and continuity. Some ancient societies are even notoriously difficult to identify as continuous entities since they have left only a neglectable amount of traces and their exact identity is a mystery (e.g. many Mesoamerican cultures). Interestingly, Mandelbaum does not mention historical agents as entities having continuous existence in time, and persons are usually studied in relation to their society (Mandelbaum 1977, 207–208).

The discussions concerning the historical realism are also reflected in the contemporary debates about historical representationalism and historical non-representationalism which have inherited many recurring issues. Comprehensive account showing the development of this debates was

¹⁷ The concept of colligatory terms is a persistent topic in analytic philosophy of history. For the most recent overview see (Kuukkanen 2015, 97–115).

¹⁸ See direct attack on Mandelbaum's distinction by (Goldstein 1986, 84–87).

provided by Eugen Zelenák (see Zelenák 2018). His overview of more radical historical representationalism follows:

According to this view, historical works should correspond to the past, if not perfectly, then at least as faithfully as possible. There should be no subjective preconceptions, no prejudicial intrusions, and no unnecessary external factors entering the process of learning about the past. Historical works should give us (relatively) direct access to the past. (Zelenák 2018, 118)

Although the idea of an historian as a scientist who is uncovering, reporting, and explaining the past may seem unproblematic at first glance, we can see that many philosophers have denounced this option. The criticism of the idea of historical realism may seem like a direct consequence of sceptical reasoning—the past is gone; therefore, we cannot know it. It does apparently stem from the fact that our knowledge of the past is limited by the available evidence which suffers from information decay over time. Consequently, since we cannot access the past directly and since we do not generally possess all necessary evidence, we cannot fully know what really happened in the real past. Thus, according to historical realism, significant portions of history as a discipline and science would be seriously restricted or even rendered impossible. However, we can clearly see that this is not the case. Historians and other scientists are successful in creating theories and narratives which have significant explanatory value, despite the fact they are not explaining the past events and entities themselves. It is actually the historical realism itself that invites scepticism about the historical methodology and their findings. Setting the real past as a touchstone for historical inquiry is to expose many (if not all) historical theories and narratives to justifiable criticism for being too speculative. In some cases, we may have overwhelming evidence at our disposal and we may be convinced that some statements about the real past are beyond doubt, however, this does not hold for a vast amount of historical narratives and does not account for diverse historical narratives or theories of “the same” subject. Even the partial realism (e.g. Mandelbaum’s distinction between objective general history and relative special histories) seems to share these issues. Should we abandon the historical realism, what alternatives do we have? Can we imagine history without the past? We will now explore the possible merits of historical antirealism.

3. Historical antirealism and banishment of the past

To say that the discipline of history is not interested in the past may seem counter-intuitive and even fatal to historians and philosophers. To fully appreciate historical antirealism, several issues must be clarified first:

- (1) Historical antirealism is not a negative claim about the ontological status of the real past. It only states that the real past is not a subject of historical inquiry and the ontological status of entities is not a pressing subject for historians to discuss. It is both possible to hold some kind of realism about the past and entities in the past and to be an historical antirealist, claiming that this real past is not the subject of the discipline of history. It is relieved of burdensome realistic load.
- (2) Historical antirealism is not saying that anything goes in the field of history. Historians are bound to provide explanations and these explanations must adhere to similar epistemic virtues (e.g. coherence, simplicity, scope, accuracy, etc.) as explanations in other areas of human knowledge. Historians are also limited by available data and accepted evidence. There is, of course, space for competing theories and disagreements among historians, however, this does not mean that historical inquiry is not a highly specialised form of knowledge that expands its achievements.
- (3) Historical antirealism is not limited to only some aspects of historical disciplines. It aims to encompass historical explanations, historical narratives, historical writing, historical research and its methods, historical representation, and various historical disciplines, including ancillary disciplines, and fields. It includes the history of ancient pottery as well as poetry, it includes national political histories and the history of modern philosophy. In other words, it does not accept Mandelbaum's distinction between general history and special histories and it strives to be inclusive, not exclusive.

Historical antirealism does not focus solely on the final products of historians that are intended for a wider audience. This was, according to L.J.

Goldstein, the common mistake made by both the covering law theoreticians and the narrativist philosophers of history alike. The history is not exclusively about writing and explaining. Before we can explain the French Revolution or write about the Peloponnesian War we must first establish them as historical entities. This is done by identifying relevant data, classification of evidence, source criticism, or by employing many other methods, available to the historian.

Goldstein criticised the narrativist philosophers of history and the analytic philosophers of history for not going beyond the texts and textbooks historians produce. It is true that historical texts are the most visible to non-historian consumers, however, they do not reveal what is unique about the historical enterprise as the way of knowing. He coined the distinction between the superstructure of history (the finished product, usually in narrative form, intended for layman consumers) and the infrastructure of history (methods and reasoning employed by historians in the course of their inquiry). Goldstein argued that philosophers have focused almost exclusively on the superstructure of history and ignored the infrastructure of history (Goldstein 1976, 139–82). The discipline of history is not about explaining the pre-given sets of facts about the past, it is not about framing the individual parts of chronicle into a single narrative. On the contrary, historians *constitute the historical past* (i.e. not the real past) before they can explain it or interpret it. This constitution of an historical past is an intellectual activity that is largely dependent on the evidence contemporary historians can identify and utilize, on the accepted procedures of historical inquiry, and on the current status of historical knowledge or discourse. The real past is virtually thrown out of the equation by historical antirealism and the roles of contemporary methods of scientific historiography and community of historians are stressed. Discussing Collingwood's Roman Britain and his conclusion from the solitary gravestone to the presence of the Irish colony (colony that "he called to the existence"), Goldstein states:

It is all well and good to say that Collingwood's statement is true only if there really was such a colony, but that is to say something that has no consequence for historical inquiry; it simply expresses the hope that historical past is identical with the real past. (Goldstein 1996, 334)

It is clear that the act of postulating the entire colony from a single gravestone as a piece of evidence is not simply an abstraction since it rather adds than abstracts.

Historical past resembles the antirealist conception of fictional models in science in a striking manner. It is created by the professionals to explain specified sets of present (encountered) phenomena. The Goldstein's point can be made even more illustrative if we apply it to the non-textual historical models. It is true that the stereotypical output of historical inquiry is a book or an article. However, we can easily encounter small-scale spatial models of historical cities like Prague in the 14th century. Such model is supposed to represent the highly idealised state of the specific city in the specific historical past, although it is not an exact full-scale reconstruction, neither it aims to represent everything exactly as it was in the real past since such accuracy is not attainable or even desirable. Individual building blocks of the model are fictional, abstract, and highly idealised. Nonetheless, various parts of the given model are based upon the evidence of a different kind. Some buildings may be included in the model on the basis of available written records or, usually incomplete, archaeological findings. Other structures may have survived to this day, although their appearance in the model may have been adjusted according to other relevant evidence. Provided we are admiring such model in a museum, some pieces of empirical data identified as an evidence are usually located nearby and they are basically substituting the role of footnotes in historical texts. The models of historical cities in the past can explain to us why the contemporary centre of the city is suddenly cut in half by the old fortification. At the same time, we can seek an explanation of some relations between various entities that become apparent in the model, for example, the locations of certain specialised structures (i.e. a division of city's quarters) or their relative distances. Such model does account for an available and identified evidence, however, does it really have to correspond to some particular state of the city in the real past? This suggestion, once again, expresses the hope that the historical past is identical with the real past, but it is of no consequence to historical research. Should we encounter new, previously unidentified, evidence, we are of course inclined to change the model of the city and the historical past. Nonetheless, we do so only to account for the evidence and its relation

to the model, we are not changing the real past, nor we are strengthening our correspondence to it. Historical facts change in time, while the real past does not. As Derek Turner states, discussing a similar example from the biology, in his book *Making Prehistory*:

Understood in this way, the conclusion is true, but it poses no threat to constructivism. On this first interpretation, to say that one fact is in the past relative to another is to say that that *the first fact was a fact at an earlier time, and that the second fact was a fact at a later time.* (Turner 2007, 153)

The primary aim of the historian is not to explain the past, but to constitute the model based on the present findings, and to explain the various parts of this model. Historians thus inquire into the present, not into the past. The historians do not usually doubt that Caesar was a real person since there is no overwhelming evidence implying that we should think otherwise. It is even better to say, that historians do not ponder over the existence of Caesar since that is not part of their work. Their task is to explain the substantial number of artefacts, texts, or relicts that are known to us. This explanation entails dealing with the fact that some evidence relevant to the given subject in the constituted model might be contradictory, however, this does not interfere with the possibility of producing a unified historical and critical account. Nonetheless, provided we want to produce comprehensive and detailed biography of Caesar and we seek to overcome contradictory evidence, we must accept that our overarching image of Caesar is just a useful fiction that helps us to accommodate contemporary data. In other cases, fragmented evidence seemingly referring to the same person may be disregarded as too vague and contradictory, therefore hindering the historical inquiry. Such cases can be described by historians as more legendary than historical characters. When we refer to an historical agent, we, in fact, refer to an entity inside a model (a constituted historical past). This is not the result of an historical character's demise, but of fragmentary and often contradictory nature of historical evidence that is treated by historians who are often producing vastly diverging accounts of historical agents. Diverse historiographical texts and interpretations are also portraying still living characters and it could be even said that their authors engage in debates with their contenders by putting forward pieces of evidence and arguing historically. Historical agents are evaluated

and described by the hypothetical references to their intentions and according to the impact they supposedly have on the historical past. Those are, however, later constructs.

The historians are fully aware of their dependence on the present evidence and its theory-laden identification. The following passage from the contemporary book on Ancient Greek history and on the life of Spartan kings Agis and Cleomenes III nicely illustrates the point. I believe that the passage is worth quoting at full length:

Written, documentary texts that might correct or supplement the opposed tendencies of the two principal literary sources are very thin on the ground. Numismatic and other material testimony tends in this case to illustrate and sometimes illuminate the literary picture rather than form the basis for an alternative account. This is partly because of the selective nature of the data we have. For example, the absence of archaeological corroboration of the literary picture of private affluence cannot be used to overthrow it, given the lack of finds from graves or private dwellings in Sparta. In short, the evidentiary situation is such that too often we cannot say for certain what events actually occurred or in what order, and usually we can only attempt to guess why. The immense modern bibliography on Agis and Cleomenes may suitably reflect the objective and symbolic importance of their reigns but it is inversely proportional to our sure knowledge of them. (Cartledge et al. 2005, 35–36)

The authors are aware of the fact that limited evidence underdetermines the results of their inquiry and that their conclusions are not infallible, however, there is still a substantial amount of empirical data present and it demands an explanation. Historical evidence can be explained only historically by constituting appropriate historical past (appropriate model), otherwise, it will remain unexplained.

Since the historical antirealism considers history to be an inquiry into the present world and since the historical past is a fiction explaining present phenomena, we must necessarily ask what phenomena and objects are eligible for historical explanation. During the second half of the twentieth century, historians started to inquire into a virtually unlimited number of

possible topics. Almost any aspect of human life may be subjected to the historical inquiry: the history of sexuality, history of sports, history of computing, history of comics, etc. At the same time, an historical explanation may have a very mundane form: if we want to explain e.g. a hammer to a child, we may say that it is a blunt tool that is used for driving nails into the wood. However, when we are confronted with a type of tool that is not used today in everyday life, we seek a different kind of explanation. We need to refer to the historical past (not the real past) and to the fictional people that were using this tool in the course of some activities. This process of identifying something as an historical evidence was nicely captured by Goldstein in one of his early works:

When we say that the starting point is the evidence, we mean only that the suspicion that there were events is suggested by the fact that there are present certain things which seem not to fit into the present context of culture and life: writings which most of us cannot read, coins which will buy nothing at the grocery, ruins of buildings and of entire cities, and so on. (Goldstein 1969, 176)

If we positively identify something as the relict of the inaccessible past and if we are able to fit this relict into some timeline (e.g. with the help of chemical analysis), we may then look for other instances of the same type of artefact, which would belong to other periods we use for classification of evidence. Once again it is seen that history explains the present world, not the past itself. The world would go on even without the historians. Nonetheless, without a systematic field of knowledge dedicated to the relicts and traces of the past, we would ascribe the ruins of Roman spas to the dwarves or other mythical beings. Historians would generally agree that they cannot predict the future in the same manner that speculative philosophers of history claim to do, however, they may have the ability to predict where to find some previously unidentified evidence on the basis of another piece of evidence. Upon establishing some prior hypothesis, historians often know which archives they should visit in their search for further evidence.¹⁹

¹⁹ A detailed overview and analysis of such occurrences can be found in (Murphey 2009, 40–46).

The historical past consists of fictional, abstract, and idealised entities we may call historical facts. Historical facts are not created independent of each other and hermetically sealed inside a factual chronicle²⁰ which needs explanation or retelling. On the contrary, when constituting an historical fact, historian already constitutes it in relation to other historical facts. The idea of past facts, suspended in some kind of ideal chronicle, waiting for the historian to explain them or to devise a narrative around them, is not only unattainable, but also misleading. Historical facts are definitely not atomic, historians do not simply build their stories from the sole pre-given blocks and do not explain isolated events by connecting them to others through the help of covering laws. When referring to historical facts, we actually refer to objects in models, not in reality.²¹ Only the past emerged in this process of the historical constitution can become the subject of further intellectual endeavour:

There are, of course, interesting things to be done with a past already emerged. One could explain it; or one could interpret it. And one could contemplate it in the belief that it must surely contain lessons for us that may be put to use as we seek to confront our present and effect our future. (Goldstein 1986, 83)

Nonetheless, it is true that historian approaches his subject equipped with the preliminary knowledge of the current discourse among the historians. Historians are not constituting their models from scratch and without any relation to other models currently accepted in the community of historians. Historians may seek to re-examine the phenomena accepted as evidence, they may strive to expand the accepted historical theories, or they may try to offer vastly different model meant to explain similar sets of phenomena. This discursive aspect of history as a discipline was only hinted at in the works of Goldstein, however, contemporary philosophers of history (not necessarily antirealists) are becoming progressively more aware of this

²⁰ The idea of chronicle could be considered as inert since it is described as a “mere” collection of meaningless and weltering facts. See (White H. 1975).

²¹ See a similar statement about the models in physics in (Cartwright 1984, 129).

aspect pertaining to the historical inquiry.²² As historians begin to sprout an interest in new topics like the history of women in politics, history of toys, or history of sports, new and previously unidentified (though available) pieces of evidence become evident. Once again, we may say that historical inquiry is essentially a contemporary inquiry (done by contemporary agents with the help of contemporary tools) into the present (inquiring into the present data labelled as evidence) while producing models explaining the present phenomena through the constitution of historical past. Contemporary agents who conduct this inquiry are present-day historians and they are working within the boundaries of their discipline. There are, of course, disagreements concerning certain peripheral subjects which cannot be easily resolved by calling out the evidence. This is because the evidence only underdetermines the theory or narrative in some cases. However, as Aviezer Tucker points out, there is a significant consensus among the historians about central issues. Tucker devoted a lot of thought to the task of defining the consensus as an epistemically significant factor: “Consensus in a uniquely heterogeneous, large, and uncoerced group of historians is a likely indicator of knowledge” (Tucker 2004, 39).

Historical antirealists could easily accept this statement, although they would swiftly deny that the knowledge indicated in such a way is knowledge of the past. It does rather imply that such knowledge encompasses the present artefacts and relicts and it does offer the best possible explanation of them in the form of a fictional historical past.

Disagreements between historians are about the proper treatment and assessment of the evidence. The competing models of historical pasts often explain the similar sets of present phenomena, however, the appropriate assessment of relevant evidence and its exact role in the course of the historical constitution is not a trivial task. A good example of the fluid nature of historical evidence could be the case of literary forgeries that have appeared in Europe during the 19th century that were intended to promote the patriotism and nationalism, like *The Dvůr Králové* and *Zelená Hora* manuscripts, found in 1817 and 1819 respectively. These forgeries have

²² Recently, a Finnish philosopher of history Jouni-Matti Kuukkanen has drawn attention to the discursive and argumentative side of historical endeavour. His current research focuses on the microhistorical epistemology; see (Kuukkanen 2017, 118).

depicted the heroic past of the Czech nation and provided a narrative that showed currently subdued nation as proud, civilised, and free people sharing a long and glorious history. At first, these carefully forged manuscripts had been taken as authentic relicts of the past and they did fuel Czech National Revival. However, by the 1880s, the authenticity of both documents was mostly rejected. This was achieved by analysing the grammatical and metrical structure of the text as well as by the extensive historical research and comparative studies. The forgeries thus ceased to be an evidence about the glorious past of the Czech nation and they become the evidence in the narrative about Czech National Revival illustrating the desperate strive of Czech intellectuals to codify the national identity. Forged manuscripts are still an historical evidence and historians can still utilize them to create a model of Czech National Revival, a model that explains a vast number of diverse texts (novels, poetry, textbooks, historiography, dictionaries, etc.) and artefacts from the 19th century but, on the other hand, it would be entirely inappropriate to include these texts when constituting the historical past of the 9th century Bohemia when taken at face value. There are, of course, other unresolved issues available to the historians of Czech National Revival, they can try to explain the motives behind the forgeries, they can ponder about the ethical issues involved, however, no historian of early medieval Bohemia finds this piece of evidence relevant to the contemporary research.

We can summarise historical antirealism as a thesis that the discipline of history is interested in explaining the present phenomena (texts, artefacts) that require the postulation of historical past to be fully explained. This historical past is a model or set of models created by historians based on the empirical data identified as an evidence. Historians proceed by discovering and identifying new pieces of evidence as well as by adjusting the models to account for previously unaccounted data which could be explained historically.

4. Caveats against historical antirealism

It is important to address several common caveats raised against historical antirealism. I will try to answer some of them that may have been the most obvious in the course of this study.

The crucial issue of historical antirealism is that its rejection of the real past as the touchstone for historical inquiry may seem unwarranted, radical, and unnecessary. We often feel that we know what happened in the past. We generally do not doubt that the Thirty Years' War started in 1618, we never question that Abraham Lincoln was a real man, and we would not deny the usefulness of concepts like the Middle Ages or the Renaissance. However, this very fact that historical antirealism seems so counter-intuitive to us is a result of our unreflected reliance upon the outcomes of historical inquiry. We are accustomed to the history textbooks and canonical historical narratives, but we cannot claim to know the real past. It could be even exaggerated that we are so immersed in historical narratives, that we do not see the real inquiry behind the history. Historians do not directly report the real past to us, they report the results of their interaction with evidence and of their critical thinking about the past they have constituted. Historical antirealism is not motivated by scepticism about the discipline of history. On the contrary, it is motivated by the wish to ward off this scepticism and to show history as a scientific discipline in its own right.

Proponents of historical antirealism are sometimes accused of cherry-picking the examples from the distant past that better suit their theory (see e.g. Nowell-Smith 1977, 4). I can easily see this caveat raised against this paper as well. However, the choice of such examples is motivated by the wish to provide comprehensive and illustrative examples of inquiries where the lack of evidence highlights the very importance of evidence-identification and its treatment. Even when assessing the memories of living witnesses during the interviews or even when utilising video or audio recordings, historians must remain critical and aware of historical methods. It would be a mistake to submit to the illusion that now they know all the relevant facts about the real past event in question. The surplus of contradictory evidence may even complicate matters. At the same time, the cherry-picking should not be a pressing concern, provided all the selected examples are true instances of historical inquiry and historical method. If analytic philosophers of history wish to explain the historical practice, they should consider all instances of historical inquiry equally, especially in the case that given field of history is accepted by the contemporary community of historians.

Philosophers of science could raise an objection that we are talking about the context of discovery,²³ and ignoring the more philosophically interesting context of justification. However, the context of discovery would imply that something is discovered in the process of historical inquiry and that is not the case according to the historical antirealism. The past is not discovered, the historical past is constituted in order to explain the data discovered before or in the course of an historical inquiry. It would be more appropriate to talk about the context of a constitution. The context of discovery more likely refers to the discovery of artefacts or texts themselves.

The last pressing matter, I would like to address is that by giving up the real past, we are potentially rejecting the value of history as a discipline. If the historical past is fiction, how can we derive any lessons from it? Why should we conduct the historical inquiry?

Apart from the fact that we often derive lessons from fictions and fables²⁴ and that many philosophers warned us not to look for some hidden wisdom in history, I would like to claim, that history is even more valuable for historical antirealism since it does explain our world here and now. It is true, that we cannot generally use this fictional knowledge to predict the future or to utilise the findings practically, but we still gain some non-trivial knowledge about the world unattainable in any other way. History can be thus valued as the discipline whose subject is not dead in the past, but it is still present in our time. It does not necessarily follow, that without the history we would be committed to making the same mistakes as our ancestors, but without it, we would have certainly lived in the world where many artefacts and texts would be unaccounted for and our knowledge of the world would be seriously lacking. Even if the historical past that historians present to us is a fictional past, it is still modelled upon the relevant empirical evidence and this process is guided by the highly systematic procedures of the historical discipline. Moreover, the outlined version of historical antirealism is much akin to the philosophical pragmatism since it is concerned with the actual historical practice, the methods of inquiry, and the

²³ A similar claim is raised in relation to Paul Roth's non-representationalism by Zeleňák (2018, 125).

²⁴ See Cartwright's distinction between the fables and their morals in comparison to models (Cartwright 1999, 36–40).

gathering of evidence. Pragmatist considerations are currently being explored by contemporary philosophers of history as they are inquiring into the discourse of professional historians.²⁵

Although the discipline of history is naturally understood as an inquiry into the past, it proves to be difficult to maintain the direct link between the contemporary historians and the real event they are supposed to examine, explain, or simply describe. Rather, historians are producing fictional models that are meant to explain present empirical data that would be otherwise inexplicable without the constitution of fictional entities. These entities involve not only abstractions, generalisations, or classifications of historical periods but also the historical agents and events themselves. Historical realism does suffer from the dangers of sceptical objections and criticism. Historical antirealism, though counter-intuitive at first sight, allows historians to retain their competences, to employ various methods, including experimental archaeology or oral history, to produce a broad range of representations, and to successfully pursue various topics through diverse historical subdisciplines. History is shown as an essential field of scientific knowledge that proves its value as an important type of inquiry into our contemporary world. Even though we may consider historical interpretations, theories, or narratives fictional, they are still useful fictions and without them, we would be robbed of large parts of knowledge we possess about the world.

Funding

The research and the paper are supported by the scientific project IRP201820 “*The Construction of the Other in Medieval Europe*” (IRP University of Ostrava).

References

- Ankersmit, Frank. 1986. “The Dilemma of Contemporary Anglo-Saxon Philosophy of History.” *History and Theory* 25 (4): 1–27. <https://doi.org/10.2307/2505129>
- Cartledge, Paul, and Antony Spawforth. 2005. *Hellenistic and Roman Sparta*. London: Routledge. <https://doi.org/10.4324/9780203482186>

²⁵ See especially (Kuukkanen 2017) and his paper with the expressive title “Moving Deeper into Rational Pragmatism” or (Fay 2017).

- Cartwright, Nancy. 1984. *How the Laws of Physics Lie*. Oxford: Clarendon Press.
<https://doi.org/10.1093/0198247044.001.0001>
- Cartwright, Nancy. 1999. *The Dappled World: A Study of the Boundaries of Science*. New York: Cambridge University Press.
<https://doi.org/10.1017/cbo9781139167093>
- Derek, Turner. 2007. *Making Prehistory: Historical Science and the Scientific Realism Debate*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/cbo9780511487385.007>
- Fay, Brian. 2017. "From Narrativism to Pragmatism." *Journal of the Philosophy of History* 11 (1): 11–21. <https://doi.org/10.1163/18722636-12341355>
- Goldstein, Leon J. 1962. "Evidence and Events in History." *Philosophy of Science* 29 (2): 175–94. <https://doi.org/10.1086/287860>
- Goldstein, Leon J. 1976. *Historical Knowing*. Austin and London: University of Texas Press.
- Goldstein, Leon J. 1986. "Impediments to Epistemology in the Philosophy of History." *History and Theory* 25 (4): 82–100. <https://doi.org/10.2307/2505133>
- Goldstein, Leon J. 1996. *The What and the Why of History: Philosophical Essays*. Leiden: Brill.
- Guérout, Martial. 1969. "The History of Philosophy as a Philosophical Problem." *The Monist* 53 (4): 563–87. <https://doi.org/10.5840/monist196953438>
- Hartmann, Stephan, Carl Hoefer, and Luc Bovens, eds. 2008. *Nancy Cartwright's Philosophy of Science*. New York: Routledge.
<https://doi.org/10.4324/9780203895467>
- Hempel, Carl G. 1942. "The Function of General Laws in History." *The Journal of Philosophy* 39 (2): 35–48. <https://doi.org/10.2307/2017635>
- Hull, David L. 1975. "Central Subjects and Historical Narratives." *History and Theory* 14 (3): 253–74. <https://doi.org/10.2307/2504863>
- Hull, David L. 1979. "In Defense of Presentism." *History and Theory* 18 (1): 1–15.
<https://doi.org/10.2307/2504668>
- Hull, David L., and Michael Ruse, eds. 2007. *The Cambridge Companion to the Philosophy of Biology*. Cambridge: Cambridge University Press.
<https://doi.org/10.1017/ccol9780521851282>
- Kuukkanen, Jouni-Matti. 2015. *Postnarrativist Philosophy of Historiography*. London: Palgrave Macmillan. <https://doi.org/10.1057/9781137409874>
- Kuukkanen, Jouni-Matti. 2017. "Moving Deeper into Rational Pragmatism." *Journal of the Philosophy of History* 11 (1): 83–118.
<https://doi.org/10.1163/18722636-12341362>
- Lewontin, Richard. 1991. "Facts and the Factitious in Natural Sciences." *Critical Inquiry* 18 (1): 140–53. <https://doi.org/10.1086/448627>

- Mandelbaum, Maurice. 1961. "Historical Explanation: The Problem of Covering Laws." *History and Theory* 1 (3): 229–42. <https://doi.org/10.2307/2504314>
- Mandelbaum, Maurice. 1977. *The Anatomy of Historical Knowledge*. Baltimore: The John Hopkins University Press. <https://doi.org/10.2307/1856346>
- Murphey, Murray G. 1986. "Explanation, Causes, and Covering Laws." *History and Theory* 25 (4): 43–57. <https://doi.org/10.2307/2505131>
- Murphey, Murray G. 2009. *Truth and History*. New York: State University of New York Press.
- Nielsen, Margit H. 1981. "Re-Enactment and Reconstruction in Collingwood's Philosophy of History." *History and Theory* 20 (1): 1–31. <https://doi.org/10.2307/2504642>
- Nowell-Smith, Patrick H. 1977. "The Constructionist Theory of History." *History and Theory* 16 (4): 1–28. <https://doi.org/10.2307/2504805>
- O'Sullivan, Luke. 2006. "Leon Goldstein and the Epistemology of Historical Knowing." *History and Theory* 45 (2): 204–28. <https://doi.org/10.1111/j.1468-2303.2006.00357.x>
- Rosenberg, Alexander. 2000. "Reductionism in a Historical Science." *Philosophy of Science* 68 (2): 135–63. <https://doi.org/10.1002/0470854189.ch7>
- Rosenberg, Alexander. 2007. "Reductionism (and Antireductionism) in Biology." In *The Cambridge Companion to the Philosophy of Biology*, edited by David L. Hull and Michael Ruse, 120–38. Cambridge: Cambridge University Press. <https://doi.org/10.1017/ccol9780521851282.007>
- Roth, Paul A. 2012. "The Pasts." *History and Theory* 51 (3): 313–39. <https://doi.org/10.1111/j.1468-2303.2012.00630.x>
- Roth, Paul A. 2017. "Essentially Narrative Explanations." *Studies in History and Philosophy of Science* 62: 42–50. <https://doi.org/10.1016/j.shpsa.2017.03.008>
- Tucker, Aviezer. 2001. "The Future of the Philosophy of Historiography." *History and Theory* 40 (1): 37–56. <https://doi.org/10.1111/0018-2656.00151>
- Tucker, Aviezer. 2004. *Our Knowledge of the Past: A Philosophy of Historiography*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/cbo9780511498381>
- Tucker, Aviezer, ed. 2009. *A Companion to the Philosophy of History and Historiography*. Oxford: Blackwell Publishing. <https://doi.org/10.1002/9781444304916>
- Tucker, Aviezer. 2011. "Historical Science, Over- and Underdetermined: A Study of Darwin's Inference of Origins." *British Journal for the Philosophy of Science* 62 (4): 805–29. <https://doi.org/10.1093/bjps/axr012>
- White, Hayden. 1973. *Metahistory: The Historical Imagination in Nineteenth-Century Europe*. Baltimore: The Johns Hopkins University Press.
- White, Hayden. 1975. "Historicism, History, and the Figurative Imagination." *History and Theory* 14 (4): 48–67. <https://doi.org/10.2307/2504665>

-
- White, Morton. 2005. *From a Philosophical Point of View*. Princeton: Princeton University Press. <https://doi.org/10.1515/9781400826469>
- Wilkins, John S. 2009. "Darwin." In *A Companion to the Philosophy of History and Historiography*, edited by Aviezer Tucker, 404–15. Oxford: Blackwell Publishing. <https://doi.org/10.1002/9781444304916.ch36>
- Wright, Georg H. von. 1971. *Explanation and Understanding*. London: Routledge & Kegan Paul.
- Zeleńák, Eugen. 2018. "Non-Representationalism in Philosophy of History: A Case Study." In *Towards a Revival of Analytical Philosophy of History*, edited by Krzysztof Brzechczyn, 116–29. Leiden: Brill. https://doi.org/10.1163/9789004356900_009

Contents

EDITORIALS

Matteo PASCUCCI and Ádám Tamas TUBOLY: <i>Preface</i>	3/318–322
Daniel von WACHTER: <i>Preface</i>	1/2–4

RESEARCH ARTICLES

Thomas ATKINSON, Daniel HILL and Stephen MCLEOD: <i>On a Supposed Puzzle Concerning Modality and Existence</i>	3/446–473
Ansgar BECKERMANN: <i>Active Doings and the Principle of the Causal Closure of the Physical World</i>	1/122–140
Sezen BEKTAŞ: <i>Knowledge after the End of Nature: A Critical Approach to Allen’s Concept of Artifactuality</i>	2/249–264
Radim BĚLOHRAD: <i>Animalism and the Vagueness of Composition</i>	2/207–227
Ralf B. BERGMANN: <i>Does Divine Intervention Violate Laws of Nature?</i>	1/86–103
David ČERNÍN: <i>Historical Antirealism and the Past as a Fictional Model</i>	4/635–659
Max CRESSWELL: <i>Modal Logic before Kripke</i>	3/323–339
Orli DAHAN: <i>There IS a Question of Physicalism</i>	4/542–571
Matej DROBŇÁK: <i>Do We Share a Language? Communitarism and Its Challenges</i>	4/572–596
Michael ESFELD: <i>Why Determinism in Physics Has No Implications for Free Will</i>	1/62–85
Adam GREIF: <i>The Morality of Euthanasia</i>	4/612–634
Lloyd HUMBERSTONE: <i>Semantics without Toil? Brady and Rush Meet Halldén</i>	3/340–404
Jeremiah Joven JOAQUIN: <i>Prospects for Experimental Philosophical Logic</i>	2/265–286
Bartosz KALUZIŃSKI: <i>Genuinely Constitutive Rules</i>	4/597–611

Robert A. LARMER: <i>The Many Inadequate Justifications of Methodological Naturalism</i>	1/5–24
Edwin MARES and Francesco PAOLI: <i>C. I. Lewis, E. J. Nelson, and the Modern Origins of Connexive Logic</i>	3/405–426
David B. MARTENS: <i>Wiredu contra Lewis on the Right Modal Logic</i>	3/474–490
Genoveva MARTÍ and José MARTÍNEZ-FERNÁNDEZ: <i>On ‘actually’ and ‘dthat’: Truth-conditional Differences in Possible Worlds Semantics</i>	3/491–504
Uwe MEIXNER: <i>Elements of a Theory of Nonphysical Agents in the Physical World</i>	1/104–121
Thomas PINK: <i>Freedom, Power and Causation</i>	1/141–168
Claudio PIZZI: <i>Alternative Axiomatizations of the Conditional System VC</i>	3/427–445
Daniel RÖNNEDAL: <i>Semantic Tableau Versions of Some Normal Modal Systems with Propositional Quantifiers</i>	3/505–536
Mirco SAMBROTTA: <i>Categories and the Language of Metaphysics</i> ...	2/186–206
Erdinç SAYAN: <i>Casting a Shadow on Lewis’s Theory of Causation</i>	2/287–297
Richard SWINBURNE: <i>The Implausibility of the Causal Closure of the Physical</i>	1/25–39
Daniel von WACHTER: <i>The Principle of the Causal Openness of the Physical</i>	1/40–61
Andrzej WALESZCZYŃSKI, Michał OBIDZIŃSKI and Julia REJEWSKA: <i>The Significance of the Relationship between Main Effects and Side Effects for Understanding the Knobe Effect</i>	2/228–248

DISCUSSION NOTES

Michael ESFELD: <i>The Principle of Causal Completeness: Reply to Daniel von Wachter</i>	1/169–174
Miloš KOSTEREC: <i>Contradiction of Modal Modification</i>	2/298–300
Daniel von WACHTER: <i>Do the Laws of Nature Entail Causal Closure? Response to Michael Esfeld</i>	1/175–184

BOOK REVIEWS

- Shih-Hsun CHEN: Martin Smith, *Between Probability and Certainty: What Justifies Belief* 2/301–305
- Pavol HARDOŠ: Lee McIntyre, *Post-Truth* 2/311–316
- Ádám Tamas TUBOLY: Willard Van Orman Quine, *The Significance of the New Logic* 2/306–310

REPORTS

- Martin VACEK: *Modal Metaphysics: Issues on the (Im)possible VII* 3/537–539