Contents

Research Articles

Discussion Note

Organon F 29 (3) 2022: 290–326
https://doi.org/10.31577/orgf.2022.29301

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

An Inferentialist Account of Fictional Names

Byeong D. Lee*

Received: 2 July 2021 / Revised: 12 October 2021 / Accepted: 1 April 2022

Abstract: The goal of this paper is to present and defend an inferentialist account of the meaning of fictional names on the basis of Sellars-Brandom's inferentialist semantics and a Brandomian anaphoric theory of reference. On this inferentialist account, the meaning of a fictional name is constituted by the relevant language norms which provide the correctness conditions for its use. In addition, the Brandomian anaphoric theory of reference allows us to understand reference in terms of anaphoric word-word relations, rather than substantial word-world relations. In this paper I argue that this inferentialist account has many important merits over its rival theories. One important merit is that it explains why we can use fictional names to make true statements, even if they lack bearers. As a consequence, this theory allows us to use fictional names without committing ourselves to an implausible ontology of fictional entities. Another important merit is that it provides a uniform semantic account of fictional names across different types of statements in which fictional names are involved.

Keywords: Fictional names; inferentialist semantics; the anaphoric theory of reference; Sellars; Brandom.

* Sungkyunkwan University
https://orcid.org/0000-0001-7080-6702
Department of Philosophy, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, Republic of Korea
bydlee@skku.edu

1. Introductory remarks

Overall, there are two approaches to the semantics of natural language. One is the *truth-conditional* approach to meaning, and the other is the *inferentialist* approach to meaning. On the former approach, the meaning of an expression is to be explained in terms of its *truth conditions*. On the latter approach, the meaning of an expression is to be explained in terms of its *inferential use*.

My goal in this paper is to present and defend a new semantic account of fictional names along the latter approach. In particular, my account of fictional names is based on Sellars-Brandom's inferentialist semantics and a Brandomian anaphoric theory of reference. On this account, the meaning of a fictional name is constituted by the relevant language norms which determine its correct use; and reference is to be explained in terms of anaphoric word-word relations, rather than substantial word-world relations. In this paper I argue that this account has many important merits over its rival theories, especially because it satisfies the following four desiderata:

- First, it is desirable to regard fictional names as genuine names rather than disguised descriptions.
- Second, it is desirable to regard fictional names as meaningful.
- Third, it is desirable to avoid attributing a bearer to a fictional name.
- Fourth, it is desirable to provide a uniform semantic account of fictional names across different types of statements in which fictional names are involved.

Thus, one important merit of this account is that it explains why we can use fictional names to make true statements, even if they lack bearers. As a consequence, this account allows us to use fictional names without committing ourselves to an implausible ontology of fictional entities. Another important merit is that it provides a uniform semantic account of fictional names across different types of statements related to fictional names.

This paper proceeds as follows. In section 2, I explain the main motivations for the aforementioned four desiderata. In section 3, I briefly explain Sellars-Brandom's inferentialist semantics and Brandom's anaphoric approach to reference. In section 4, I explain how fictional and non-fictional

names are different in their language norms. Finally, in section 5, I discuss the merits of my inferentialist account.

2. Four desiderata

In this section, let me explain the main motivations for the aforementioned four desiderata for a semantic theory of fictional names.

To begin with, consider the following two sentences:

Bertrand Russell smokes.
Sherlock Holmes smokes.

These two sentences share the same form 'x smokes', where 'x' is a place holder for a name. As Adams et al (1997, 131) point out, fictional names such as 'Sherlock Holmes' play the role of a name syntactically, and also in inferences. For example, from the premise that Sherlock Holmes is not married, we can infer that he has no wife. Moreover, when one reads a sentence like 'Sherlock Holmes is a detective', it is natural to imagine that the property of being a detective is attributed to someone. One important alternative is the descriptivist view of names, which holds that ordinary proper names are semantically equivalent to definite descriptions. Notably, Currie (1990) argues for a descriptivist theory of fictional names, according to which the meaning of a fictional name can be understood in terms of the entire set of descriptions associated with the name. But such a descriptivist theory is vulnerable to well-known problems pointed out by Kripke (1980).

One important problem arises from Currie's claim that the meaning of a fictional name is equivalent to the entire set of descriptions associated with the name. What should be noted in this regard is that one can successfully use a fictional name, even if one knows very little about those descriptions. For example, one can assert that Sherlock Holmes is a detective, even if one knows very little about this fictional character. For these reasons, fictional names should be regarded as genuine names, if possible. This is my first desideratum.

Let us move on to the second desideratum. The most influential semantic view of proper names today is referentialism, according to which the

semantic contribution of a name is its referent.[1] A strong version of referentialism is Millianism, the view that the semantic content of a name is exhausted by its referent. But this view faces a serious challenge with regard to fictional names. For fictional names such as 'Sherlock Holmes' seem to lack bearers, and hence this view seems to imply that they are devoid of semantic content. One option that referentialists can take is to bite the bullet by admitting that fictional names are not meaningful. Notably, according to Walton's make-believe theory (1990), when we are engaging with a fictional story, we are just pretending that fictional names are meaningful.

But we have no real difficulty in understanding sentences containing fictional names, such as 'Sherlock Holmes is a detective'. If this is the case, fictional names contained in such sentences are better understood as being meaningful. In addition, as Salmon (2005, 76) points out, pretend use is not real use, and so by merely pretending that a name like 'Sherlock Holmes' has a particular use, no real use is attached to the name. In other words, a pretend use of a name does not generate a real name. What is noteworthy in this regard is that we are free to pretend whatever we like. But it is not correct to say, for example, that Sherlock Holmes is a ballet dancer. This indicates that 'Sherlock Holmes' as a real name has the correctness conditions for its use. But if 'Sherlock Holmes' were not meaningful, we could hardly say that this name has the correctness conditions for its use. Therefore, *pace* the pretense view, it is desirable to retain the natural opinion that fictional names are meaningful. This is my second desideratum.

Another option for referentialists is to hold that fictional names have bearers. But this option requires them to accept the realist view which states that our reality includes fictional objects as bearers of fictional names. Two popular realist approaches are Meinongianism (e.g., Parsons 1980; Routley 1980; Zalta 1983, 1988) and Artifactualism (e.g., Salmon 2005). By appealing to a metaphysical distinction between 'there is' and 'exists', Meinongians hold that there are such things as fictional objects, but those things are non-existent objects. By contrast, according to Artifactualism, fictional objects are abstract artifacts which are created by human practices,

[1] This view is associated with philosophers such as Kripke (1980), Donnellan (1974), and Kaplan (1979).

and those abstract entities actually exist, because artifactualists do not accept the distinction between 'there is' and 'exists'.

By contrast, the anti-realist view denies that our reality contains such fictional objects. It seems that there are empty names, which lack bearers. For example, consider names from myths and mistaken scientific theories such as 'Zeus', Pegasus', and 'Vulcan'. It also seems that we can make true negative existential statements such as 'Vulcan does not exist.' And to say that Vulcan does not exist seems tantamount to saying that 'Vulcan' has no bearer. If, as these examples suggest, there are indeed empty but meaningful names, then fictional names like 'Sherlock Holmes' can also be meaningful, even if they lack bearers. Besides, the alleged fictional objects, whether they are non-existent or abstract, are at least metaphysically controversial objects. Moreover, semantics should not meddle with our naturalistic world-view, if possible. For these reasons, it would be worth exploring the view that fictional names can be meaningful, even if they lack bearers. This is my third desideratum.[2]

Finally, my fourth desideratum is that it is desirable to provide a uniform semantic account of fictional names across different types of statements in which fictional names are involved. To illustrate, consider the following three kinds of statements.

(1) Sherlock Holmes is a detective.
(2) According to The Hound of the Baskervilles, Sherlock Holmes is a detective.
(3) Sherlock Holmes is a fictional character created by Conan Doyle.

We can distinguish between discourse that is *internal* to a fiction and discourse that is *external* to the fiction. In internal discourse, a fictional statement like (1) is to be understood from the perspective within fiction. By contrast, in external discourse, a metafictional statement like (3) is to be understood from the perspective of the real world, outside fiction. To put

[2] Here I do not mean to suggest that there is no such account yet. One notable example is Sainsbury's Fregean view (2005; 2010). This view allows us to understand fictional names as empty but meaningful. But he takes a truth-conditional approach to meaning. By contrast, as noted, the goal of this paper is to offer an alternative account along the inferentialist approach to meaning.

the point another way, the name 'Sherlock Holmes' in (1) is used *fictionally*, that is, in internal discourse. By contrast, the name 'Sherlock Holmes' in (3) is used *metafictionally*, that is, in external discourse. Note that Sherlock Holmes is understood as a flesh and blood individual in internal discourse, whereas Sherlock Holmes is understood as a fictional character in external discourse.[3]

But according to Recanati (2018), there is the third type of use for fictional names: *parafictional uses*. On his view, there is no such person as Sherlock Holmes in the world, and so 'Sherlock Holmes' in (1) is an empty name. As a consequence, when a fictional name is used fictionally, it is not genuinely referential. By contrast, according to Recanati, when a fictional name, as used in (3), is used metafictionally, it refers to a cultural artifact. Thus, metafictional uses of a fictional name are genuinely referential. The question then is how to understand statements like (2), i.e., statements about what is true in some fiction but which are not part of the original storytelling. On Recanati's view, like metafictional statements, parafictional statements such as (2) are true or false; and they are to be evaluated from the perspective outside fiction. On the other hand, like fictional statements, the properties which parafictional statements ascribe to the putative referent of a fictional name are the kind of properties which fictional statements ascribe, that is, properties such as *being a detective* and *playing the violin*. Note that these are properties suitable for flesh and blood individuals, not for abstract objects. Along these lines, Recanati argues that parafictional statements share features with both fictional and metafictional statements.

At this point, an important question arises: Is there a uniform semantic analysis of fictional, parafictional, and metafictional statements? What is noteworthy in this regard is that we seem to run into trouble when we try to provide a uniform semantic account of fictional names used in these different types of statements. Let me illustrate this point. Suppose that a realist approach is true, so that a fictional name like 'Sherlock Holmes' refers to an abstract object across these different types of statements. Then we face a problem in understanding fictional statements like (1), because properties like *being a detective* are not suitable for abstract objects. On the

[3] For a more detailed discussion of the internal/external distinction, see Semeijn & Zalta (2021, 172–75).

other hand, suppose that a fictional name like 'Sherlock Holmes' refers to a flesh and blood individual, so that properties like *being a detective* can be attributed to Sherlock Holmes. But then we run into a problem in understanding metafictional statements like (3), because flesh and blood individuals are not the kind of things that can be created by a novelist. Certainly, it is desirable to avoid this kind of problem.[4] This is my fourth and last desideratum.

Two cautionary remarks might be in order here. The first is concerned with the above-discussed desiderata. In this paper, I will not provide any further defense for them. This is not because it needs no more defense, but rather because a proper defense of these desiderata would take me too far from the main goal of this paper, which is to present an inferentialist account of fictional names, rather than criticizing its rival theories. Besides, I think these desiderata are reasonable, so that it is worthwhile to explore an account which satisfy all of them.

The second cautionary remark is also related to the goal of this paper. My inferentialist account of fictional names is deeply indebted to Brandom's works. Unfortunately, however, he has not provided a separate account of fictional names. But on my view fictional names such as 'Sherlock Holmes' lack bearers. And whether or not a name has a bearer affects its meaning. As a consequence, whether an expression is used as a fictional name or as non-fictional name makes a significant difference to its meaning. This is why we need a separate account of fictional names. In addition, I will also mention some additional differences between Brandom's own account and my account in due course.[5]

[4] Due to this kind of problem, Semeijn & Zalta (2021) argue that a uniform semantic treatment of fictional names is required across fictional, parafictional and metafictional discourse. They also argue that their object theory can provide such a uniform semantic treatment. But their object theory does not meet my third desideratum, namely that it is desirable to avoid attributing a bearer to a fictional name.

[5] See especially footnotes 12 and 17.

3. Sellars-Brandom's inferentialist semantics and Brandom's anaphoric approach to reference

My inferentialist account of fictional names is based not only on Sellars-Brandom's inferentialist semantics, but also on Brandom's anaphoric theory of reference, although my account differs from their views in some respects. Thus, let me briefly explain these theories in this section.

According to Sellars-Brandom's inferentialist semantics, we can understand the meaning of a linguistic expression in terms of language norms (or rules) that bind those who use the expression. In particular, according to Sellars (1963, 327-331), there are three kinds of language norms. The first is language-entry norms. The circumstances in which an expression is correctly applied are an essential aspect to the meaning of the expression, and such circumstances can be non-linguistic. And language-entry norms prescribe linguistic moves in response to such non-linguistic circumstances. For example, in the presence of a visibly red thing, one is allowed to assert 'This is red', and one is prohibited to assert 'This is blue'. The second is language-language norms. These norms are concerned with the appropriate consequences of application of an expression. Thus these norms prescribe linguistic moves in response to linguistic episodes. For example, under the circumstances in which one can assert 'This is red', one is allowed to infer 'This is colored', but one is prohibited to infer 'This is blue'. The third is language-exit norms. These norms prescribe non-linguistic moves in response to linguistic episodes. For example, under normal circumstances and barring a change of mind, a person should pick up a red apple after he says, 'I'll pick up a red apple'. On Sellars's view, it is these three kinds of language norms for an expression that are constitutive of the meaning of the expression. As I will argue in the next section, we can understand the meaning of fictional names in a similar way.

Let us now turn to Brandom's anaphoric theory of reference. To begin with, it is important to note the distinction between a bottom-up approach and a top-down approach. Consider the following simple sentence: 'Lassie is a dog.' This sentence consists of two components. One is the name 'Lassie', and the other is the predicate 'is a dog'. On a bottom-up approach, we first need to explain the meanings of such sub-sentential expressions and

then, on the basis of these meanings, we should explain the meanings of sentences constructed by such sub-sentential expressions, and finally proprieties of inferences in which those sentences are involved. Accordingly, on this bottom-up approach, the meaning of a name should be intelligible independently of the meaning of a sentence in which it occurs. Contemporary representational approaches to semantics typically adopt this approach. One notable example is Tarskian model-theoretic semantics.[6] But Sellars-Brandom's inferentialist denies this kind of bottom-up approach. As mentioned before, this semantics explains the meaning of an expression in terms of its *inferential use*. What should be noted in this regard is this: It is *sentences* that can play the basic inferential roles of premises and conclusion in inferences. Accordingly, the meanings of sub-sentential expressions such as singular terms and predicates have to be projected from the inferential roles of sentences. Therefore, according to the inferentialist semantics, the meanings of sub-sentential expressions are not prior to the meanings of sentences in which those sub-sentential expressions occur. To put the point another way, the inferentialist semantics adopts a top-down approach. This top-down approach starts by explaining the inferential relations between sentential claims, and then explains the meanings of sub-sentential expressions in terms of their potential contribution to those inferential relations.

With this difference in mind, consider the following statement:

(4) 'China' refers to China.

On the traditional, non-deflationary approach to reference, 'refers' in (4) expresses a substantial relation between a name and its referent as an extra-linguistic entity, and so (4) is true because the name 'China' stands in a *substantial referential relation* to a certain extra-linguistic entity, namely the world's third largest country (by land area). Let me briefly explain why this view is problematic.

The first thing to note is that it is certainly possible for someone to use an expression type 'China' as a name of something other than the country, for example, as the name of his pet dog. Thus, if a person successfully uses an expression token 'China' as the name of the country, then this is not because the expression type 'China' stands in a referential relation to the

[6] For example, see Montague (1974) and Tarski (1983).

country, but because s/he *uses* the expression token 'China' in a certain way. If so, how can we distinguish between cases in which it is used as the name of the country and cases in which it is used as the name of a particular pet dog. One plausible answer is this: A person uses 'China' as the name of the country, if s/he uses it in accordance with a language norm such as that one may (or ought to) apply the name only to the particular country. If one instead uses 'China' in accordance with a different language norm such as that one may apply the name only to a certain pet dog, then it is used as the name of a pet dog. If this is correct, we cannot determine what is referred to by such an expression independently of the relevant language norms. But as noted before, on the bottom-up approach, the meaning of a name should be intelligible independently of the meaning of a sentence in which it occurs. In particular, according to the direct reference theory of names, names are *directly referential*. For example, the name 'Lassie' means Lassie because the former directly refers to the latter. What is meant here by 'directly referential' is that the name contributes nothing but its referent to the meaning of a sentence in which it occurs, and the referential relation between the name and its referent has priority over the meaning of the whole sentence, so that the meaning of 'Lassie' can be understood independently of the meaning of the whole sentence.[7]

At this point, it is important to recognize that the correctness conditions for the use of a name can hardly be established independently of any sentence in which it occurs. Let me explain. Suppose that Jones has a pet dog called 'Lassie'. Under what conditions can we say that the one referred to as 'Lassie' is indeed Jones's pet dog? As pointed out before, we cannot determine what is referred to by such a name independently of the relevant language norms, such as that one may apply the name only to Jones's pet dog. How then can this kind of language norm be established? First, Jones's pet dog must be given this name 'Lassie'. And this naming process typically involves using sentences such as 'Let us call my pet dog Lassie'. Without using such a sentence, we can hardly conduct this kind of naming ceremony. Second, this kind of language norm must be maintained socially by virtue of positive and negative social sanctions. For example, if someone misuses

[7] Millians such as Donnellan (1974), Kaplan (1975), and Salmon (1986; 2005) uphold this view.

this name to talk about something other than Jones's pet dog, this misuse should be corrected by using sentences such as 'That thing is not Lassie' or 'Lassie is the name of Jones's pet dog'. If these considerations are correct, there are no such things as *directly referential* relations between our language and the world.

Of course, this does not mean that there are no relationships between tokens of 'China' and the world's third largest country. Admittedly, there might be some correlations between those tokens of 'China' and the world's third largest country. Nonetheless, those correlations would be very complex. More importantly, there is no good reason to pick out any of those correlations as *the* desired word-world referential relation. In this regard, it is important to note that meaning is *normative*. For example, it is our language norm that we may apply the name 'China' only to a particular country. Given this language norm, you are *allowed* to use 'China' to talk about the particular country, and you can be *subject to criticism* for making a linguistic error, if you use it to talk about another country such as Japan. In this sense, the use of an expression has a normative implication. But any factual correlation by itself does not have this kind of normative implication.

Along the above lines, Brandom argues that it is very difficult to explain what the aforementioned word-world relation really is, and so we had better pursue an alternative approach to reference. On his anaphoric theory, reference is not a substantial concept. For example, the quoted name 'China' in (4) is *mentioned* and the name appearing after the expression 'refers' is *used*; and 'refers' here does not express a substantial relation between a linguistic expression and an extra-linguistic entity. Instead, 'refers' here expresses an *anaphoric word-word relation* between the mentioned name and the used name. What then is an anaphoric word-word relation? Consider the following statement:

If *Mary* wants to leave on time, *she* should leave now.

In this conditional statement, 'she' is a pronoun that is used instead of the proper name 'Mary' in the antecedent. Consequently, the token of 'she' bears an anaphoric word-word relation with the token of 'Mary'. And such an anaphoric relation is a *commitment-preserving* link in the following sense: If anyone treats two word tokens as anaphorically related, then s/he

is thereby committed to treating both as having the same inferential significance. Brandom argues that we should understand the notion of reference in the same way. Suppose that someone named 'Joe' makes the following statement to another person with the name 'Jim':

> I should have known better than to let the mechanic Binkley work on my car. That airhead misadjusted the valves.

Suppose also that Jim forgot the name 'Binkley', but he nonetheless remembers that Joe called the mechanic as 'that airhead'. Then he may say:

> For car repair, don't go to the mechanic Joe referred to as 'that airhead'.

According to Brandom (1994, 305; 2005b, 265–66), in this discourse, the description 'the mechanic Joe referred to as "that airhead"' is a lexically complex pronoun that takes the token of 'that airhead' originally used by Joe as its anaphoric antecedent. Brandom calls such a description 'an anaphorically indirect definite description'. If the antecedent and dependent tokens are sufficiently close to each other in time, space, or audience attention, one may use lexically simple pronouns such as 'he', 'she' or 'it'. As for distant antecedents, however, one might be required to use such indirect definite descriptions, which give us more information about their antecedents. On Brandom's view, we can understand the expression 'refers' as a pronoun-forming operator that is used to form such an anaphorically indirect definite description. Accordingly, in the above case, the token of 'the mechanic Joe referred to as "that airhead"' bears an anaphoric word-word relation with the token of 'that airhead', and such an anaphoric relation is a commitment-preserving link; that is, since Jim uses the former instead of the latter, if Jim is committed to holding that the one Joe referred to as 'that airhead' is *F*, then he should also be committed to holding that the one he referred to as 'the mechanic Joe referred to as "that airhead"' is *F*. On the anaphoric theory of reference, therefore, we can understand the expression 'refers', not in terms of a substantial relation between a linguistic expression and an extra-linguistic entity, but rather in terms of an anaphoric word-word relation.[8]

[8] The anaphoric theory of reference is a deflationary theory of reference, rather than a semantic theory of names. And there are many reasons for the inferentialist

At this point, it would be worthwhile to compare the anaphoric theory with the causal-historical theory of reference. To begin, direct referentialists sharply distinguish the semantics of names from what Lycan (2008, 52) calls 'a philosophical theory of referring'. The semantics of names explains what names contribute to the meanings of sentences in which they occur; for such a semantic account, Kripke (1980) proposes a Millian theory of names, according to which the sole semantic contribution of a name is its bearer. By contrast, a philosophical theory of referring addresses the question of how one's utterance of a name is tied to the object that gets referred to by that utterance; for this account, Kripke proposes the causal-historical theory of reference. On this theory, roughly, a speaker, using a name on a particular occasion, refers to an object if there is a causal-historical chain of reference-preserving link leading from the speaker's use of the name on that occasion ultimately to the event of the object's being given that name. What should be noted about the causal-historical theory of reference in conjunction with the Millian theory of names is that a causal-historical chain of tokens of a name is nothing more than a *reference-preserving* link. The reason is clear. On the Millian theory, the semantic content of a name is exhausted by its referent, and hence it is not legitimate to appeal to anything other than its referent in explaining what the name contributes to the meaning of a sentence in which it occurs. In addition, on this non-deflationary theory of reference, if a dependent token of a name like 'Joe Biden' inherits a referent from an antecedent token, this is because the dependent token is anaphorically related to some antecedent token which stands in a substantial referential relation to a certain extra-linguistic entity, namely Joe Biden himself.

---

to take a deflationary approach to reference. The most important reason is this: On Sellars-Brandom's inferentialist semantics, the meaning of an expression should be explained in terms of a substantial notion of inference, rather than in terms of alleged substantial truth-theoretic notions. In other words, this semantics denies that truth conditions should play the fundamental role in semantics that the truth-conditional semantics give to them. As a consequence, on this semantics, its basic notion is correct inference, rather than truth-theoretic notions such as truth and reference. If, however, the notion of reference is substantial to the effect that inference depends on reference, we can hardly understand the notion of inference independently of the notion of reference.

As Brandom (2005a, 248) points out, the considerations that motivate the causal-historical theory can be understood equally as motiving the anaphoric theory. But there are still important differences between these two theories of reference. First, on the anaphoric approach, an anaphoric relation is a *commitment-preserving* link, rather than a reference-preserving link, so that if anyone treats two word tokens as anaphorically related, then s/he is thereby committed to treating both as having the same inferential significance. Second, on the anaphoric theory, the sameness of reference is achieved by an anaphoric relation, but reference here should be understood as a deflationary notion. That is, we should understand the expression 'refers', not in terms of a substantial relation between a linguistic expression and an extra-linguistic entity, but rather in terms of an anaphoric word-word relation.

Here again, a cautionary note might be necessary. It is beyond the scope of this paper to fully defend the anaphoric theory of reference. See Brandom (1994; 2005b) for a detailed defense of this theory. My goal is instead to defend an inferentialist account of fictional names by assuming that it is a viable theory of reference. And based on the anaphoric theory, as we will see, we can explain the meaningfulness of fictional names, even if these names lack bearers.

4. The differences between fictional and non-fictional names in language norms

As mentioned in the previous section, on Sellars-Brandom's inferentialist semantics, the meaning of an expression is constituted by the relevant language norms which determine its correct use. In this section, I argue that the meaning of fictional names can be understood in a similar way. For example, I argue that the meaning of 'Sherlock Holmes' is constituted by the relevant language norms which determine its correct use. In addition, I argue that the main differences in meaning between fictional and non-fictional names arise from the fact that the uses of these names are governed by different language norms.

Let us start by considering the following question: How can we distinguish works of fiction from works of non-fiction? According to Walton (1990), non-fiction invites belief, whereas fiction invites imagining without belief, and so works of fiction are distinguished from works of non-fiction in that the former essentially involve a proposal or invitation to imagine. Besides Walton, several writers such as Currie (1990), Lamarque & Olsen (1994), Davies (2007), Stock (2011), and García-Carpintero (2013) have defended various versions of the prescriptions to imagine account of fiction. Among these versions, I agree with García-Carpintero's version on which the norms of fiction are constitutive of fictional discourse. On this normative version, if one engages with a fictional story, one is thereby prescribed to imagine as the story says, and such a prescription has normative force for the audience; in other words, the audience is subject to the following norm: 'Imagine that *p* if, according to the story, *p*.' For example, when one engages with a Holmes story such as *The Hound of the Baskervilles*, one is prescribed to imagine that Sherlock Holmes is a detective. And refusing to imagine in this way is tantamount to refusing to participate in a human practice of consuming such a story as a work of fiction.[9]

What is important to note here is that the prescriptions to imagine account of fiction is an account for the distinction between fictional and non-fictional works, but not an account for the meaning of fictional names. For example, Walton subscribes to the direct reference theory of names. Thus, on his view, fictional names are devoid of semantic content, because they lack bearers. And this is why, as mentioned in section 2, he advocates a pretense view, according to which when we are engaging with a fictional story, we are just pretending that fictional names are meaningful. Likewise, García-Carpintero's normative account of fiction-making is an account for the distinction between fiction and non-fiction. On this account, the norms

[9] This does not mean that we have an absolute obligation to imagine whatever is said in a work of fiction. As García-Carpintero (2013, 346) points out, such prescriptions can be understood as weak directives such as proposals or invitations to imagine. Thus, when one is invited to imagine as a fictional story says, one can refuse the invitation. Nonetheless, insofar as one engages with a fictional story by (implicitly) accepting this invitation, one is thereby prescribed to imagine as the story says.

of fiction are constitutive of fictional discourse, but those norms have no direct bearing on the meaning of fictional names. Therefore, the prescriptions to imagine account of fiction does not force us to accept a particular meaning theory of fictional names. My proposal is to combine a version of this account with the inferentialist semantics.

To begin, my inferentialist account of fictional names adopts a Brandomian anaphoric theory of reference. As discussed in the previous section, we can understand the expression 'refers', not in terms of a substantial relation between a linguistic expression and an extra-linguistic entity, but rather in terms of an anaphoric word-word relation. And an anaphoric chain of tokens of a name is a *commitment-preserving* link, so that if anyone treats two tokens of a name as anaphorically related, then s/he is thereby committed to treating both as having the same inferential significance. The same points apply to fictional names. But one important difference is this: Even if a fictional name lacks a bearer, a token of the name can initiate an anaphoric chain, which can be continued by other tokens of the same name or tokens of a pronoun. For example, one can make up a story in the following way:

> Sherlock Holmes is a detective. He lives with Dr. Watson in Baker Street. He is interested in Early English Chapters, and so on.

Here a token of the name 'Sherlock Holmes' initiates an anaphoric chain that is continued by tokens of the pronoun 'he'. It is (partly) by virtue of such an anaphoric chain that the correctness conditions for the use of 'Sherlock Holmes' are established. And it is also by virtue of such an anaphoric chain that different tokens of a fictional name are all about the same fictional character.

In addition, on my inferentialist account, we can also understand the meaning of a fictional name in terms of the relevant language norms which determine its correct use. But there are still important differences between fictional and non-fictional names in their language norms. The most important difference is that, in the case of a fictional name such as 'Sherlock Holmes', the correctness conditions for its use depend on a work of fiction in which it occurs, whereas this is not the case for non-fictional names such as 'Joe Biden'. Let me elaborate on this point.

First, an author can make up a story by constructing anaphoric chains of fictional names in the way suggested above, even if those names lack bearers. Second, if a work of fiction has successfully been published, certain norms of fiction are thereby established, and those norms are such that, if one engages with a fictional story, one is thereby prescribed to imagine as the story says; in other words, one is subject to the following norm: 'Imagine that p if, according to the story, p.'[10] Third, we can understand such norms as providing the correctness conditions for the use of fictional names. For example, insofar as we engage with a Holmes story such as *The Hound of the Baskervilles*, it is correct to imagine that Sherlock Holmes is a detective, whereas it is not correct to imagine that he is a ballet dancer. Note that the Holmes story does not say, nor imply, that Sherlock Holmes is a ballet dancer. Along these lines, we can argue that the correctness conditions for the use of a fictional name depend on a work of fiction in which it occurs. And since an author can make up a story by constructing anaphoric chains of fictional names, even if those names lack bearers, the meaningfulness of a fictional name does not depend on the condition that it has a bearer. By contrast, non-fictional names like 'Joe Biden' have bearers, and the correctness conditions for the use of a non-fictional name does not depend on any work of fiction. For example, we can say that Joe Biden is the 46th president of the United States. Clearly, this correct use of the name does not depend on any work of fiction.

An important related point is that, as mentioned in section 2, a fictional name can be used *fictionally*, that is, in internal discourse, and it can also be used *metafictionally*, that is, in external discourse. For example, 'Sherlock Holmes' can be used in a fictional statement such as 'Sherlock Holmes is a detective', and it can also be used in a metafictional statement such as 'Sherlock Holmes is a fictional character'. At this point, it is important to note that the reason why fictional names like 'Sherlock Holmes' can be correctly used not only in internal discourse, but also in external discourse is that authors such as Conan Doyle have successfully introduced the relevant fictional names into our language by having written fictional stories

[10] Note that at least under normal circumstances a fictional name like 'Sherlock Holmes' becomes a part of a public language only after a work of fiction containing the name has been published. In other words, before a work of fiction is published, there is no public meaning for fictional names which are introduced in the fiction.

containing them. But again, this is not the case for non-fictional names. For example, we can correctly say that Joe Biden is president. But this statement is neither fictional nor metafictional. This is because its correctness has nothing to do with any work of fiction.

The differences between fictional and non-fictional names in their language norms are discussed in more detail below.

4.1. Language-entry norms

As mentioned before, the meaningfulness of a fictional name does not depend on the condition that it has a bearer. As a consequence, fictional and non-fictional names are bound to be different with regard to language-entry norms, especially when fictional names are used metafictionally. Let me explain.

A fictional name like 'Sherlock Holmes' is introduced by a work of fiction. Thus, unlike the case of a non-fictional name, the original anaphoric chain of tokens of such a fictional name is constructed by an author (or authors), even if it lacks a bearer. Once such an anaphoric chain of tokens of the name is thus constructed, an anaphoric relation holds among those tokens, and it is by virtue of the anaphoric relation that those tokens of the fictional name are all about the same fictional character. Therefore, we may say that a fictional character is constructed (partly) by virtue of an anaphoric chain of tokens of a fictional name constructed by an author. And the meaning of a fictional name like 'Sherlock Holmes' is constituted by the relevant language norms which determine its correct use. And it is due to such correctness conditions that 'Sherlock Holmes' can be meaningfully used, despite the fact that it lacks a bearer. For this reason, a fictional name like 'Sherlock Holmes' has no language-entry norm that allows us to say 'Here is Sherlock Holmes' in the presence of the bearer of the name. And the fact that a fictional name lacks such a language-entry norm is no defect in its meaning at all. This is one important difference between fictional and non-fictional names in language-entry norms.[11]

[11] Here I do not mean to deny that sentences like 'Here is Sherlock Holmes' might be used in response to some non-linguistic circumstances, when this fictional name is used fictionally. For example, it is possible that an actor A_1 who is playing the

There is also a related difference between fictional and non-fictional names. As mentioned in the previous section, the meaning of a non-fictional name is constituted in part by its language-entry norms, which prescribe allowable linguistic moves in response to non-linguistic circumstances. For example, in the presence of Jones, one is allowed to say 'Here is Jones'. Note that the name 'Jones' ought to be applied only to Jones. And this language norm requires settling non-linguistic circumstances in which this name is correctly used. But there is no such requirement for the use of a fictional name. For fictional names lack bearers.

A cautionary remark might be necessary. Here I am not denying that one can write a fictional story containing non-fictional names. For example, many non-fictional names including 'Napoleon Bonaparte' appear in Leo Tolstoy's *War and Peace*, because this historical fiction chronicles the history of the French invasion of Russia and the impact of the Napoleon era on Tsarist Russia. In addition, our understanding of such a historical novel relies partly on our knowledge related to the historical background of such a novel. The question then is how we should understand the meaning of a non-fictional name used in a work of fiction. Let us focus on the use of 'Napoleon' in *War and Peace*. When we read such a historical novel, in the absence of contrary indications, we are supposed to understand non-fictional names like 'Napoleon' in an ordinary way. For example, in the absence of contrary indications, 'Napoleon' is used as the name of a real person, who was one of the greatest military commanders in history, Emperor of France, and one who invaded Russia.

But it should be noted that a historical fiction is a fiction, not a history book. Thus, a fiction author is not prevented from writing a fictional story which is not true of a historical figure such as Napoleon. As has been emphasized, insofar as we engage with a fictional story, we are prescribed to imagine as the story says. Non-fictional names are not exceptions. As a consequence, the Napoleon character in *War and Peace* does not have to be

role of Dr. Watson in a Sherlock Holmes movie utters a sentence 'Here is Sherlock Holmes' in the presence of another actor A_2 who is playing the role of Sherlock Holmes. But we should not forget that A_1's statement should be understood from the perspective of the fiction. Thus, in such a case, we are prescribed to imagine that Watson says 'Here is Sherlock Holmes' in the presence of Sherlock Holmes.

the same as the real-life character of Napoleon. Therefore, although 'Napoleon' is the name of a real person, when it is used in the fiction, its use is governed by the following norm: 'Imagine that *p* if, according to the story, *p*.' As a consequence, the name 'Napoleon' is governed by different language norms, depending on whether it is used in a fictional context or in a non-fictional context. This means that whether it is used in a fictional context or in a non-fictional context makes a difference to its meaning. In this regard, it might be worth considering a historical movie in which an actor, say A_1, utters the sentence 'That's Napoleon', pointing to another actor, say A_2, who is playing the role of Napoleon. In this case, we may say that the name 'Napoleon' is correctly applied to A_2. This is because A_1 uses the name 'Napoleon' in a fictional context, and also because we as movie watchers are prescribed to imagine that A_2 is Napoleon. To put the point another way, the reason why the name 'Napoleon' is correctly applied to A_2 in this case is that its use is governed by the norms of fiction. Another thing to note is that the Napoleon character in a fiction can be very different from the real-life character of Napoleon if many things the fiction says about Napoleon are not true of the real Napoleon.

4-2. Language-language norms

Let us now turn to language-language norms. Compare the following two modal claims:

(5) Joe Biden might not have been president.
(6) Joe Biden might not have been a person.

There is an important sense in which we can make such a *de re* modal claim as (5), but we can hardly make such a *de re* modal claim as (6). Let me explain. The non-fictional name 'Joe Biden' is currently used as the name of a real person in our language, presumably by virtue of the fact that his parents gave the name to him. Thus, with regard to such a non-fictional name, we can, in principle, do the following: By pointing to a certain real person having a specific origin, and saying 'This one is Joe Biden', we can fix that person as the bearer of this name, and then start describing various hypothetical scenarios, continuing to use 'Joe Biden' as the name of the same person. This is why we can easily think about the possibility that Joe

Biden is not president, while continuing to use 'Joe Biden' as the name of the same person. This is also why we can say that 'Joe Biden is president' is contingently true, or equivalently that being president is an accidental property of Joe Biden. But insofar as we are using 'Joe Biden' as the name of the same individual, we can hardly describe a hypothetical scenario in which Joe Biden is not a person. Note that if someone uses 'Joe Biden' as a name of something other than a person, we can hardly interpret him as talking about the same individual whom we are talking about by using the name 'Joe Biden'. For this reason, we can say that there is no possible world in which Joe Biden is not a person, or equivalently that being a person is an essential property of Joe Biden.

We can also explain the above difference between (5) and (6) by virtue of the difference between inferences based on matters of fact and inferences based on language norms. Compare the following two inferences:

(5') 'x is Joe Biden' → 'x is president'.
(6') 'x is Joe Biden' → 'x is a person'.

Given the fact that we currently use 'Joe Biden' as the name of a certain person, we ought to use the name in accordance with (6'). This is the case even when we consider various hypothetical scenarios about Joe Biden. Recall that there is no possible world in which Joe Biden is not a person. And this is the reason why we can hardly make such a *de re* modal claim as (6). Along these lines, we can argue that (6') is a meaning-constitutive inference. By contrast, we can easily describe a possible scenario in which Joe Biden is not president. In this connection, I agree with Sellars (1948) that the meaning of an expression is constituted only by *counterfactually robust inferences*. If this is correct, (5') is not meaning-constitutive.[12]

[12] There are two approaches to inferentialist semantics. On Sellars's inegalitarian view, the meaning of an expression is constituted only by counterfactually robust inferences. By contrast, on Brandom's egalitarian view, all inferences including ones based on ancillary information are meaning-constitutive (see Brandom 1994, 634; 2010, 168). It is beyond the scope of this paper to settle whose view is correct. Thus, let me just mention two important reasons why I prefer Sellars's view. First, on this view, there is a principled distinction between inferences directly relevant to meaning and inferences based on ancillary information, and hence we can preserve our natural

This is not analogous to fictional names. Compare the following two statements:

(5'') Joe Biden is president.
(1) Sherlock Holmes is a detective.

As previously pointed out, we can correctly say (5''), but it is neither fictional nor metafictional, because its correctness has nothing to do with any work of fiction. In addition, since Joe Biden is a real individual whom we can locate in our physical world, we can fix that individual as the bearer of the name 'Joe Biden', and then start describing various hypothetical scenarios, continuing to use it as the name of the same individual. Therefore, (5'') is contingently true, and so (5') is not meaning-constitutive. By contrast, (1) is a fictional statement. And there are good reasons to think that the following is meaning-constitutive.

(1'') 'x is Sherlock Holmes' → 'x is a detective'.

First, an author can construct a fictional character in such a way that the nature of the fictional character is completely determined by the author.

Second, an author can also construct a fictional character in such a way that the distinction between necessity, possibility, and impossibility is ignored. For example, as in *The Metamorphosis* written by Franz Kafka, if an author wants, he can write a story in which a human being gets transformed into a non-human creature. This means that any property which an author has ascribed to a fictional character is partly constitutive of the fictional character. In this regard, it is worth recalling this: If a work of fiction has successfully been published, certain norms of fiction are thereby

intuition that the validity of such inferences as that 'Lassie is a dog' → 'Lassie is an animal' has a direct bearing on the meaning of an expression involved in such inferences, whereas this is not so with such inferences as that 'Lassie is a pet dog → 'Lassie is adorable'. Second, we can provide a good explanation as to why we don't usually feel much pressure for meaning instability. For example, we learn the meaning of 'dog' by learning to use it in accordance with the relevant public language norms such as that 'x is a dog' → 'x is an animal'. And we can share the meaning of 'dog' because we are bound by those same norms. Note that this intersubjective role of language norms is secured by the high stability of counterfactually robust inferences, not by ancillary information which can differ from person to person.

established, and those norms are such that, if one engages with a fictional story, one is thereby prescribed to imagine as the story says.

Third, strictly speaking, there is no cross-work identity for fictional characters. Let me explain. It seems that a fictional character appearing in a certain story can reappear in other stories. For example, Sherlock Holmes appearing in *The Hound of the Baskervilles* apparently reappears in other stories such as *A Samba for Sherlock* written by Brazilian author Jô Soares. There are many similarities between the Holmes character in *The Hound of the Baskervilles* and the Holmes character in *A Samba for Sherlock*. For example, these two characters share many properties such as *having the same name*, *being a detective*, and *having Dr. Watson as a friend*. But there are also a number of dissimilarities. For example, unlike the former case, we are prescribed to imagine that Soares's Holmes is fumbling and nearsighted, and so he fails to solve the crimes he has undertaken to investigate.

The question then is whether or not these two are the same fictional character. On my view, these two are, strictly speaking, different fictional characters, because some norms of fiction to which we are subject in each of these cases are different. For example, as noted, we are prescribed to imagine that Soares's Holmes is fumbling and nearsighted. It is also important to observe that, for any predicate Conan Doyle ascribed to Sherlock Holmes in *The Hound of the Baskervilles*, a new author can write a new fictional story in such a way that the predicate is not ascribed to a fictional character with the same name. This shows that there are no essential properties that make two characters in different stories one and the same character. Along these lines, we may argue that, strictly speaking, there is no cross-work identity for fictional characters.

If so, what kind of relation holds between the Holmes character in *The Hound of the Baskervilles* and the Holmes character in *A Samba for Sherlock*? According to Sellars (1974), our concept (or meaning) can undergo a change. For example, the concept of mass changed during the transition from Newtonian mechanics to relativistic mechanics. In this case, the Einsteinian concept of mass is not simply other than the Newtonian concept of mass; for Newtonian mass and Einsteinian mass are so functionally similar that they can be regarded as varieties of mass. Along these lines, Sellars argues that these two concepts are closely related counterpart concepts.

And it is due to this counterpart concept relation that we may say that the concept of mass underwent a change from Newtonian mass to Einsteinian mass, rather than saying that an old concept was simply replaced by a wholly different concept. On my view, this Sellarsian view of counterpart concepts could be extended in an analogous manner to cases where the name of a fictional character reappears in a different work of fiction. Then we can say that the Holmes character in *The Hound of the Baskervilles* bears a kind of counterpart relation to the Holmes character in *A Samba for Sherlock*, although, strictly speaking, there is no cross-work identity between these two. And essentially the same point applies to fictional characters belonging to a group of stories written by the same author. Note again that, for any predicate ascribed to a fictional character by an author in a novel, the same author can in principle write a new fictional story in such a way that the predicate is not ascribed to a fictional character with the same name. Here I do not mean to deny that the audience could regard fictional characters with the same name which appear in a series of works written by the same author as the same fictional character in a loose sense.[13] But again, it needs to be emphasized that there are no essential properties that make two characters in different stories one and the same character in a strict sense.

If the above considerations are correct, one important difference between fictional and non-fictional names in language-language norms can be

[13] On my account, strictly speaking, there is no cross-work identity for fictional characters. In connection with this claim, an anonymous reviewer made the following suggestion: "The choice of taking a novel as the unit that determines a certain fictional character seems arbitrary: why not chapter/ paragraph/ sentence? I would suggest just taking 'story/narrative' as a primitive notion and allowing that a narrative can be spread out over several books. *Lord of the Rings* for instance consists of a couple of books but it is one narrative (and this entire narrative determines identity conditions for fictional characters)." This suggestion can be accommodated in my account. On my account, if a work of fiction has successfully been published, certain norms of fiction are thereby established, and those norms are such that, if one engages with a fictional story, one is thereby prescribed to imagine as the story says; in other words, one is subject to the following norm: 'Imagine that p if, according to the story, p.' This view is compatible with the claim that the story/narrative can be spread out over several books.

illustrated by the fact that (5') is not meaning-constitutive, whereas (1'') is meaning-constitutive. And this view provides a good explanation as to why it is difficult to make a *de re* modal claim about a fictional character such as 'Sherlock Holmes might not have been a detective'. First, with regard to a fictional name such as 'Sherlock Holmes', we cannot do the following: By pointing to a certain fictional character around us, and saying 'This one is Sherlock Holmes', we first fix that fictional character as the bearer of this name, and then start describing various hypothetical scenarios, continuing to use 'Sherlock Holmes' as the name of the same fictional character. This is because there is no real object that we can fix as the bearer of this name at the very beginning. Second, the above kind of modal claim presupposes that a fictional character could lose some of its properties without losing its identity. But fictional characters are not real agents who are capable of making free choices, and the nature of a fictional character is completely determined by the author (or authors). More importantly, our reality does not contain fictional characters as real objects. In this regard, it should be recalled that fictional names such as 'Sherlock Holmes' lack bearers.

There is an additional difference between fictional and non-fictional names in language-language norms. One can use 'Sherlock Holmes' in a meta-afictional context. In such a case it should be used (or understood) partly in accordance with the following language-language norm: 'x is Sherlock Holmes' → 'x is a fictional character'.

Here it might be worth considering one possible objection. As pointed out before, on my inferentialist account, we can make *de re* modal claim about real individuals, but we cannot make *de re* modal claim about fictional characters. But this view seems to conflict with Friend's claim about *counter-fictional imagining* (see Friend 2011). For example, according to Kafka's *The Metamorphosis*, Gregor Samsa wakes up one morning to find himself transformed into a beetle-like creature. But on Friend's view, we can imagine *counter-fictionally* that Gregor Samsa was transformed into a cockroach-like creature instead of a beetle-like creature. Let me address this problem. Consider the following modal claim:

(7) Gregor Samsa might have been transformed into a cockroach-like creature instead of a beetle-like creature.

Admittedly, we can imagine a fictional situation such that Gregor Samsa was transformed into a cockroach-like creature. But imagining something differs from making a modal claim. What should be noted in this regard is that imagining is independent of truth and belief. For we can imagine not only what is false but also what is metaphysically (or physically) impossible. To put the point another way, one can imagine anything as one pleases, without committing oneself to holding what is imagined. For example, one can imagine that one goes faster than the speed of light, that one goes back to the past, or that one is transformed into a therianthrope. In addition, to say that one can write a fictional story in which a person is transformed into a non-human is tantamount to saying that one can imagine such a scenario. And as Kafka's *The Metamorphosis* illustrates, someone can certainly write such a fictional story. One more thing worth mentioning in this connection is that one's imagination does not have to be strict or complete. For these reasons, when one imagines a fictional situation which can be described partly by using the name of an object, the constraint of its identity conditions can be loosened. This is why one can imagine a fictional situation such that Joe Biden is transformed into a non-human, even if being a person is an essential property of Joe Biden.

Keeping the above point in mind, consider the claim that Gregor Samsa might have been transformed into a cockroach-like creature. And let us call the cockroach-like creature Gregor Samsa*. Then, a question arises regarding whether Gregor Samsa* is really the same fictional character as Gregor Samsa we know from Kafka's story. On my view, the answer is 'No'. Let me explain.

Insofar as we are engaging with Kafka's story, we are not prescribed to imagine that Gregor Samsa was transformed into a cockroach-like creature. In addition, from the perspective outside the fictional story, the only solid fact we have is that Kafka wrote a certain story of Gregor Samsa. Therefore, what we can say here is just a possibility that Kafka might have written a somewhat different story using the name 'Gregor Samsa'. But if such were the case, strictly speaking, we would have a different story, and so a different fictional character. Of course, I do not mean to deny that one can imagine contrary to what Kafka's story of Gregor Samsa says. But this does not show that (7) is a true modal claim about the Gregor Samsa character.

Recall that what one can imagine is not constrained by the conditions for the identity of an object. The question then is: How can we make sense of the alleged counter-fictional situation that Gregor Samsa was transformed into a cockroach-like creature?

The aforementioned view of counterpart concepts can help us on this matter as well. As pointed out before, there are no essential properties that make two characters in different stories one and the same character in a strict sense. Thus, we can say, for example, that the Holmes character in *The Hound of the Baskervilles* bears a kind of counterpart relation to the Holmes character in *A Samba for Sherlock*. And this view could be applied to alleged counter-fictional imagining. As mentioned before, our concept can undergo a change. Thus, we can think about the possibility that our current concept of Gregor Samsa would undergo a change. For example, if we imagined that Gregor Samsa was transformed into a cockroach-like creature, our present concept of Gregor Samsa would undergo a change. Along these lines, we may argue that our imagination about Gregor Samsa is based on our current concept of Gregor Samsa, but nonetheless the former is not constrained by the latter. Recall again that what one can imagine is not constrained by the conditions for the identity of an object. We may also argue that our present concept of Gregor Samsa bears a counterpart relation with the concept of Gregor Samsa*, and so when we imagine that Gregor Samsa was transformed into a cockroach-like creature, what we are really imagining is a fictional situation which can be described partly by using a counterpart concept of Gregor Samsa. Hence, on my view, to say that one can imagine counter-fictionally that Gregor Samsa was transformed into a cockroach-like creature is tantamount to saying that one can imagine a very similar fictional situation which can be described partly by using a counterpart concept of Gregor Samsa.[14]

[14] At this point, it is worth considering an objection raised by an anonymous reviewer. On my account, there is no cross-work identity for fictional characters, and an anaphoric relation is a commitment-preserving link (at least insofar as an anaphor and its antecedent are used in the same context). If so, how can this account square with the fact that anaphoric pronouns can be used across different works of fiction. For example, consider the following sentences:

4.3. Language-exit norms

Finally, fictional and non-fictional names also have some important differences in language-exit norms, which prescribe non-linguistic moves in response to linguistic episodes.

(i) In Doyle's stories, Holmes is a successful detective, but in *A Samba for Sherlock*, he is a failure.

(ii) According to the Harry Potter books, Harry is a straight guy, but in a certain fan-fiction, he is bisexual.

In the first sentence, the name 'Holmes' and the pronoun 'he' are anaphorically related. In the second sentence, the name 'Harry' and the pronoun 'he' are also anaphorically related. But we should notice that sometimes a pronoun can be used for a usage where there is an imprecise match between a pronoun and its antecedent. To illustrate, consider the following two sentences:

(iii) Jane wears her hat almost every day, but Susan wears it only on special event days.

(iv) According to Newtonian mechanics, the mass of an object is constant, but according to Einstein's theory of relativity it is interconvertible with energy.

The pronoun 'it' contained in (iii) is used instead of its antecedent 'her hat' to avoid unnecessary repetition. But it and its antecedent are not co-referential. A similar point can be made about (iv). The pronoun 'it' here is also used to avoid unnecessary repetition. But the meaning of 'mass' is different between Newtonian mechanics and Einstein's theory of relativity. As these examples illustrate, when an anaphor and its antecedent are used across different contexts, it is not required that they mean the same thing.

With the above point in mind, consider sentences (i) and (ii) again. On my account, we can regard the first conjunct of (i) as reporting a fact about what Doyle's stories say, and the second conjunct of (i) as reporting a fact about what Soares's novel entitled 'A Samba for Sherlock' says. In a similar vein, we can regard the first conjunct of (ii) as reporting a fact about what Harry Potter books say, and the second conjunct of (ii) as reporting a fact about what a certain fan-fiction related to the Harry Potter books says. In addition, the pronoun 'he' both in (i) and (ii) can be best interpreted as being used as a pronoun of laziness. Therefore, cases of this kind do not pose a serious problem for my account. For cases of this kind are compatible with my claim that, strictly speaking, there is no cross-work identity for fictional characters. What is also noteworthy in this connection is that on my account, Harry Potter in the Harry Potter books bears a kind of counterpart relation to Harry Potter in the fan-fiction.

Let us begin by considering the case in which fictional names are used in internal discourse. As has been emphasized, when we engage with a fictional story, we are prescribed to imagine as the story says. As a consequence, fictional names are not governed by the typical kind of language-exit norms determining the correct use of non-fictional names. For example, if you engage with a Holmes story such as *The Hound of the Baskervilles*, you are prescribed to imagine that Sherlock Holmes is a detective. But you are not prescribed to make any non-linguistic move in response to such a fictional sentence. Therefore, fictional statements do not play the kind of conduct-guiding role that non-fictional statements typically have.[15]

And when fictional names are used in external discourse, these names and non-fictional names are bound to have some important differences in language-exit norms, because fictional names lack bearers. There are two kinds of cases in which a non-linguistic move is prescribed in response to a linguistic episode. In the first kind of case, a speaker's statement can prescribe a hearer to make a non-linguistic move. For example, in the case of the name of a real person, a speaker can ask the bearer of the name to do something (e.g., closing a door). In the second kind of case, a speaker's intention statement can prescribe the speaker herself to make a non-linguistic move. For example, a person can ask a question to another person named 'Smith' after she says, 'I'll ask a question to Smith'. But fictional characters are not real agents who can do something in response to a speaker's request, and so there are no language-exit norms for a fictional name by which a fictional character is prescribed to make a non-linguistic move. In addition, since fictional names like 'Sherlock Holmes' lack bearers, we are not allowed to make intention statements of the following sort: 'I will meet Sherlock Holmes', and 'I will hire Sherlock Holmes to solve this case.'

[15] Here I do not deny that there might be non-linguistic moves based on fictional discourse. For example, it is possible that an actor A_1 in a Sherlock Holmes movie utters a sentence 'I will meet Sherlock Holmes', and then there follows a scene in which A_1 meets another actor A_2 who is playing the role of Sherlock Holmes. But we should not forget that A_1's statement should be understood from the perspective of the fiction. And if the script of the movie does not include a scene in which A_1 meets A_2, A_1's fictional statement that he will meet Sherlock Holmes does not make him act so as to meet A_2.

As argued before, a fictional name is governed by different language norms, depending on whether it is used in internal discourse or in external discourse. In this regard, it is noteworthy that one can be subject to different norms in different contexts. For example, it is possible that some veterans in a certain country are subject to civilian law in peacetime, whereas they are subject to military law in wartime. In such a case, those veterans should understand that they must follow different laws, depending on whether they are in a war situation or not. In a similar vein, there is nothing strange about the fact that when a fictional name 'Sherlock Holmes' is used in internal discourse, we are prescribed to imagine that Sherlock Holmes is a detective, whereas when this fictional name is used in external discourse, we can say that Sherlock Holmes is a fictional character. And one who fully understands the meaning of a fictional name should know how to use the name not only in internal discourse but also in external discourse. Therefore, it is due to the very nature of a fictional name that it is governed by different language norms, depending on whether it is used in internal discourse, or in external discourse.[16]

5. Merits of my inferentialist account

So far, I have defended an inferentialist account of fictional names on the basis of Sellars-Brandom's inferentialist semantics and a Brandomian anaphoric theory of reference. In this final section, let me briefly explain why my account satisfies the four desiderata discussed in section 2.

First, on my inferentialist account, fictional names are genuine names whose meanings are constituted by the relevant language norms which determine their correct use. As a consequence, this account does not face

[16] What is noteworthy in this regard is that fictional names are not ambiguous at least in the sense that terms like 'bank' are ambiguous. If someone says 'Jones owns a bank', what is said by her could mean that Jones owns a financial institution or that Jones owns the land alongside a body of water. So disambiguation is needed to understand what she said. But this is not the case for fictional names. Depending on whether a fictional name is used in internal discourse, or in external discourse, we can understand what it means, with no need for disambiguation.

problems with the descriptivist view of names. For example, this account can explain why one can successfully use a fictional name, even if one knows very little about the descriptions associated with the name. On my inferentialist account, one may (or ought to) use a fictional name in accordance with the relevant language norms. And we can engage in the social division of linguistic labor with regard to those language norms. Therefore, we can use a fictional name by deferring to authorities on the norms governing the use of the name. This is why one can successfully use fictional names, even if one knows very little about the descriptions associated with the name.

Second, fictional names are also meaningful. This is again because the meaning of a fictional name is constituted by the relevant language norms which determine its correct use. In this regard, it is noteworthy that the crucial difference between fictional and non-fictional names is not that, unlike the latter, the former are not meaningful, but rather that at least some of their language norms are different. And this should be the case because, whether or not a name has a bearer affects its relevant language norms, and fictional names such as 'Sherlock Holmes' lack bearers.

Third, my account also explains the meaning of fictional names without attributing bearers to these names. The reason is clear. On my account, the meaningfulness of a fictional name does not depend on the condition that it has a bearer. As has been emphasized, the meaning of a fictional name is constituted by the relevant language norms which determine its correct use. And even though a fictional name lacks a bearer, the anaphoric theory of reference discussed in section 3 allows us to genuinely use (or understand) the name in accordance with the relevant language norms. This is because the sameness of reference can be achieved by an anaphoric relation, and reference here should be understood in terms of an anaphoric word-word relation, instead of a substantial referential relation between a linguistic expression and an extra-linguistic entity. Therefore, we can retain the natural opinion that fictional names are indeed meaningful, even if they lack bearers.[17]

[17] Brandom (1994, especially 440-449) defends the so-called 'relaxed account of existence'. On this account, to say that o exists is to say that there is some address in some structured space of addresses to which o may be assigned. What then is a structured space of addresses? According to Brandom, there are some privileged

Fourth, and finally, my account provides a uniform semantic account of fictional names across different types of statements in which fictional names are involved. Consider the following sentences again:

(1) Sherlock Holmes is a detective.

sets of expressions playing the role of what he calls 'canonical designators'; and a disjoint class of canonical designators defines a distinct structured space of addresses at which objects may be located. On the basis of this notion of canonical designators, he explains what it is to be committed to the existence of a kind of object: Roughly, to be committed to the existence of a kind of object is to treat a certain class of designators as canonical designators. Furthermore, on his view, there are at least three distinguished classes of canonical designators, and so at least three species of existence. On this view, Sherlock Holmes has fictional existence because 'Sherlock Holmes' is a fictional canonical designator and so Sherlock Holmes has an address in a certain fictive space. Along these lines, Brandom argues that physical existence, arithmetic existence, and fictional existence are species of existence. This relaxed account of existence has some merits and demerits. But I don't have enough space to discuss them here. Thus, let me confine myself to pointing out the most important reason why I do not accept this account.

Realists about fictional characters believe that our reality contains fictional characters such as Sherlock Holmes. In other words, they are ontologically committed to such fictional objects. By contrast, anti-realists deny that our reality contains such fictional objects. Considering this important disagreement about ontological commitment, it is contentious whether our reality does contain fictional characters such as Sherlock Holmes. For this reason, it would be misleading to say that fictional characters such as Sherlock Holmes exist as objects in some structured space of addresses. Here I do not deny that Brandom uses the expression 'existence' in a relaxed manner. Nevertheless, his insistence that physical existence, arithmetic existence, and fictional existence are species of existence could easily mislead us into neglecting the aforementioned important ontological disagreement about fictional characters. Hence, at least from an anti-realist point of view, we had better bring out the important difference between merely fictional characters and ontologically real objects such as physical objects more clearly, rather than covering up them by using the term 'existence' in the relaxed manner that Brandom suggests. Besides, on my view, fictional names such as 'Sherlock Holmes' lack bearers. If this is correct, it would be very misleading to say that fictional characters such as Sherlock Holmes exist as objects in some structured space of addresses.

(2) According to *The Hound of the Baskervilles*, Sherlock Holmes is a detective.

(3) Sherlock Holmes is a fictional character created by Conan Doyle.

On my account, when a fictional name is used in internal discourse, it is governed by the following norm: 'Imagine that p if, according to the story, p.' As a consequence, we are prescribed to imagine that Sherlock Holmes is a detective. In this sense, we can take (1) to be true as a fictional statement. In addition, we can understand (1) in accordance with the relevant language norms. For example, we can understand it partly in accordance with the following language-language norm: 'Sherlock Holmes is a detective' → 'he collects information to solve crimes'. Therefore, we can explain the meaning of the fictional name 'Sherlock Holmes' here in accordance with the inferentialist semantics.[18]

Let us consider (3) before (2). It is an empirical question whether there is such a fictional character as Sherlock Holmes. This is because such a question depends on whether a relevant work of fiction has actually been written. And we know that Arthur Conan Doyle wrote some fictional stories in

[18] On the prefix view such as Brock's prefix fictionalism (2002), we cannot take such a fictional statement as 'Sherlock Holmes is a detective' as its face value, ascribing it the same subject-predicate form that a parallel description about a real-life character has; instead, we should always regard it as an abbreviation for a longer sentence beginning with a story operator such as 'according to *The Hound of the Baskervilles*'. Notice that the embedded sentence is part of a prefixed sentence, and so the longer sentence would not be meaningful unless the embedded sentence is meaningful. Therefore, unless the prefix strategy provides a plausible account of fictional names within the scope of a story operator, this strategy would only defer the task of providing a plausible semantic account of fictional names. By contrast, my inferentialist account allows us to take such a fictional sentence as 'Sherlock Holmes is a detective' as a simple subject-predicate sentence, in a similar way as we can take such a non-fictional sentence as 'Joe Biden is president' as a simple subject-predicate sentence. Instead of taking 'Sherlock Holmes is a detective' as an abbreviation for a longer sentence beginning with a story operator, my account distinguishes between two discourses in which the fictional name is used. If it is used in internal discourse, we can ascribe the predicate 'is a detective' to Sherlock Holmes, whereas if it is used in external discourse, we can ascribe the predicate 'is a fictional character' to Sherlock Holmes.

which the Holmes character is portrayed. Therefore, we can say that there is such a fictional character as Sherlock Holmes. In addition, as pointed out in the previous section, when one uses 'Sherlock Holmes' in external discourse, one must use it partly in accordance with the following language-language norm: 'x is Sherlock Holmes' → 'x is a fictional character'. For these reasons, (3) is a metafictional statement which we can endorse as true.

Finally, my inferentialist account allows us to understand (2) without difficulty. As mentioned in section 2, (2) is a parafictional statement. On my view, however, the alleged parafictional statements are a species of metafictional statements. Recall that (2) is used in external discourse. And in external discourse, we can say that (2) is true on the grounds that Conan Doyle wrote a fictional story entitled 'The Hound of the Baskervilles' in which Sherlock Holmes is a detective. In this regard, it is worth recalling that it is an empirical question whether a certain author wrote a novel in which a certain predicate is ascribed to a certain fictional character. For this reason, we may regard (2) as reporting a fact about a certain work of fiction and what it says.

In addition, my account has no difficulty in explaining how the properties which parafictional statements ascribe to the putative referent of a fictional name are the kind of properties that fictional statements ascribe. For example, they are properties like *being a detective* or *playing the violin*, that is, properties suitable for flesh and blood individuals, not for abstract entities. On my inferentialist account, we can understand the embedded sentence 'Sherlock Holmes is a detective' of (2) in accordance with the relevant language norms; and when this sentence is used within a story operator, we can understand 'Sherlock Holmes' partly in accordance with the following language-language norms: 'x is Sherlock Holmes' → 'x is a human being'; 'x is Sherlock Holmes' → 'x can die'.

Furthermore, the recent debate on anaphoric dependencies across mixed discourse does not pose a serious problem for my account.[19] For example, consider the following two statements:

(8) Sherlock Holmes is a fictional character created by Conan Doyle. In Conan Doyle's stories, he is a detective.

[19] For a detailed discussion of this debate, see Semeijn & Zalta (2021, 171-75).

(9) In *War and Peace*, Napoleon is a hero. But actually, he was nothing but a dictator.

(8) is mixed discourse in that Sherlock Holmes's being a fictional character is said from the perspective outside fiction, and his being a detective is said from the perspective within fiction. And (9) is also mixed discourse in that Napoleon's being a hero is said from the perspective within fiction, and his being a dictator is said from the perspective outside fiction.

The first thing to note is that the pronoun 'he' in (8) can be replaced by its anaphoric antecedent 'Sherlock Holmes'. For the pronoun in this case is used *in the lazy way* in order to avoid unnecessary repetition of the name. The second thing to note is that as pointed out before, my account takes the alleged parafictional statements to be a species of metafictional statements. Note that the first and second statements in (8) are to be evaluated from the perspective outside fiction. Accordingly, parafictional statements can be mixed with metafictional statements. Therefore, on my account, there is no difficulty in understanding mixed discourse such as (8).

(9) can be dealt with in a similar way. The first thing to note is that a non-fictional name like 'Napoleon' can be used in a fiction. For this reason, 'Napoleon' in the first statement of (9) and 'Napoleon' in the second statement are used as the name of the same real person. But what should be noted at this point is that the Napoleon character in a fiction can be very different from the real-life character of Napoleon if many things the fiction says about Napoleon are not true of the real Napoleon. To put the point another way, the first statement is a metafictional statement about the Napoleon character in *War and Peace*, and the second statement is a non-fictional statement about Napoleon. Therefore, both the first and second statements in (9) are to be evaluated from the perspective outside fiction. What is worth recalling here is that the alleged parafictional statements are a species of metafictional statements. Therefore, there is no difficulty in understanding (9).

One more thing to note about mixed discourse is that, as Semeijn and Zalta (2021) argue, the literature on mixed discourse only establishes a need for a uniform analysis across parafictional and metafictional statements. This fact fits very well with my account. On my account, as has been emphasized, parafictional statements are a species of metafictional statements.

Thus, there is no wonder that parafictional statements can be mixed with metafictional statements. And fictional statements can hardly be mixed with metafictional or parafictional statements. This is because we can hardly take both the perspective within fiction and the perspective outside fiction simultaneously.

To conclude, my inferentialist account satisfies the aforementioned four desiderata for a sematic account of fictional name. Hence, I argue that my account provides a viable and attractive account of the meaning of fictional names.[20]

References

Adams, Fred, Fuller, Gary, and Stecker, Robert. 1997. "The Semantics of Fictional Names." *Pacific Philosophical Quarterly* 78 (2): 128–148. https://doi.org/10.1111/1468-0114.00032

Brock, Stuart. 2002. "Fictionalism about Fictional Characters". *Noûs* 36 (1): 1–21. https://doi.org/10.1111/1468-0068.00358

Brandom, Robert. 1994. *Making It Explicit*. Cambridge: Harvard University Press.

Brandom, Robert. 2005a. "Expressive versus Explanatory Deflationism about Truth." In *Deflationary Truth*, edited by Bradley Armour-Garb and JC Beall, 237–57. Chicago and La Salle: Open Court.

Brandom, Robert. 2005b. "Reference Explained Away: Anaphoric Reference and Indirect Description." In *Deflationary Truth*, edited by Bradley Armour-Garb and JC Beall, 258–81. Chicago and La Salle: Open Court.

Brandom, Robert. 2010. "Inferentialism and Some of Its Challenges. In *Reading Brandom: On Making It Explicit*, edited by Bernhard Weiss and Jeremy Wanderer, 159–80. London: Routledge.

Currie, Gregory. 1990. *The Nature of Fiction*. Cambridge: Cambridge University Press.

Davies, David. 2007. *Aesthetics and Literature*. London: Continuum.

Donnellan, Keith. 1974. "Speaking of Nothing." *Philosophical Review* 83 (1): 3–31. https://doi.org/10.2307/2183871

Friend, Stacie. 2011. "The Great Beetle Debate: A Study in Imagining." *Philosophical Studies* 153 (2): 183–211. https://doi.org/10.1007/s11098-009-9485-4

[20] I would like to thank an anonymous reviewer for his or her valuable comments on an early version of this article.

García-Carpintero, Manuel. 2013. "Norms of Fiction-Making." *British Journal of Aesthetics* 53 (3): 339–357. https://doi.org/10.1093/aesthj/ayt021

Kaplan, David. 1979. "On the Logic of Demonstratives." *Journal of Philosophical Logic* 8 (1): 81–98. https://doi.org/10.1007/BF00258420

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

Lamarque, Peter, and Olsen, Stein Haugom. 1994. *Truth, Fiction and Literature*. Oxford: Clarendon Press.

Lycan, William. 2008. *Philosophy of Language: A Contemporary Introduction,* Second edition. London: Routledge.

Montague, Richard. 1974. *Formal Philosophy: Selected Papers of R. Montague*, edited by Richmond H. Thomason. New Haven: Yale University Press.

Parsons, Terence. 1980. *Nonexistent Objects*. New Heaven: Yale University Press.

Recanati, François. 2018. "Fictional, Metafictional, Parafictional." *Proceedings of the Aristotelian Society* 118 (1): 25–54. https://doi.org/10.1093/arisoc/aoy001

Routley, Richard. 1980. *Exploring Meinong's Jungle and Beyond*. Canberra: Australian National University.

Sainsbury, R. M. 2005. *Reference without Referents*. Oxford: Clarendon Press.

Sainsbury, R. M. 2010. *Fiction and Fictionalism*. London: Routledge.

Salmon, Nathan. 1986. *Frege's Puzzle*. Cambridge: MIT Press.

Salmon, Nathan. 2005. "Nonexistence." In his *Metaphysics, Mathematics, and Meaning*, 50–90. Oxford: Clarendon Press.

Sellars, Wilfrid. 1948. "Concepts Involving Laws and Inconceivable Without Them." *Philosophy of Science* 15 (4): 287–315. doi:10.2307/2268988

Sellars, Wilfrid. 1963. *Science, Perception and Reality*. Atascadero: Ridgeview.

Sellars, Wilfrid. 1974. "Conceptual Change." In his *Essays in Philosophy and its History*, 172–88. Dordrecht: D. Reidel.

Semeijn, Merel, and Zalta, Edward. 2021. "Revisiting the 'Wrong Kind of Object' Problem." *Organon F* 28 (1): 168–197. https://doi.org/10.31577/orgf.2021.28108

Stock, Kathleen. 2011. "Fictive Utterance and Imagining." *Aristotelian Society Supplementary Volume* 85 (1): 145–61. https://doi.org/10.1111/j.1467-8349.2011.00200.x

Tarski, Alfred. 1983. "The Concept of Truth in Formalized Language." In *Logic, Semantics, and Metamathematics* (Second Edition), translated by J. H. Woodger, and edited by John Corcoran, 152–278 Oxford: Clarendon Press.

Walton, Kendall. 1990. *Mimesis as Make-Believe*. Cambridge: Harvard University press.

Zalta, Edward N. (1983). *Abstract Objects*. Dordrecht: Dordrecht Reidel.

Zalta, Edward N. (1988). *Intentional Logic and the Metaphysics of Intentionality*. Cambridge: The MIT Press.

Organon F 29 (3) 2022: 327–347
https://doi.org/10.31577/orgf.2022.29302

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Pragmatization of Narrative in Wittgenstein's Later Philosophy: A Modern Perspective

Andrii Synytsia*

Received: 20 December 2020 / Revised: 21 December 2021 / Accepted: 22 July 2022

Abstract: The main views on the nature of narrative in Wittgenstein's later philosophy are analyzed. It is shown how, realizing the research narrative, he paid attention to the linguistic means of expression of our thoughts, reference to the actual reality, the logical component of argumentation. It is shown that in order to place worldview accents more clearly and strengthen the expressive effect of thought, Wittgenstein pragmatized the narrative, in particular, used metaphors, images of learning, took into account the historico-cultural context. It is important for him to show that the form of the narrative influences what meanings the interlocutor will comprehend. Through a system of micro- and macro-narratives, Wittgenstein intended to express his opinion as clearly as possible, although he made the reader an active participant in the narrative. The thinker did not deviate from the analytico-scientific standards of philosophizing, although he showed that the relevant analysis of the narrative is significantly complicated by the ambiguities of its interpretation, the uniqueness of human experience and the identity of each narrator's value system. It is argued that a pragmatic approach to narrative analysis significantly expands the research methodology of the analytic thinker and, accordingly, makes it possible to deepen our understanding of reality

* Ivan Franko National University of Lviv
https://orcid.org/0000-0002-0983-7187
Department of History of Philosophy, Faculty of Philosophy, Ivan Franko National University of Lviv, Universytetska str. 1, 79000 Lviv, Ukraine.
andrii.synytsia.edu@gmail.com

and human existence, as well as more clearly define the specifics of their knowledge.

Keywords: Narrative; pragmatic analysis; form of life; language; Wittgenstein.

1. Introduction

Research methodology in different philosophical traditions significantly influences the way of knowledge representation. This is especially noticeable in analytic philosophy. The emphasis on the analysis of language, the intention not to start research without clarifying the meaning of key terms, special interest in arguments, refusal to use intuitive, metaphysical or non-empirical experience—these are the important points that determined the narrative of analytic thinkers. In the beginning, this state of affairs significantly limited their ways of presenting arguments. As a result, socio-political issues remained out of their attention for some time. Narratives that appealed not to emotions but to reason could appear to be a sophistication of language, a set of unsubstantial and uninteresting propositions (cf. Unger 2014).

However, already in Wittgenstein's later philosophy there was an attempt to significantly modify the analytic methodology and in some way (in a way to pragmatize research narratives) to solve the challenges it faced. To demonstrate how this happened, it is appropriate to consider Wittgenstein's later work and analyze his narratives, which, under the influence of interest in the pragmatics of language, increasingly acquired plot completeness and originality. It is important to show that chosen by him pragmatic way of interpreting the narrative (which, in the end, was not always developed in analytic philosophy) makes it possible to successfully avoid such accusations as those mentioned above.

2. Theoretico-methodological foundations of narration

First of all, it is necessary to consider the fact that there was a so-called anthropological turn in Wittgenstein's later philosophy (Gebauer 2017) and

it is quite natural that in his researches various pragmatic (Boncompagni 2016; Garrison 2017), phenomenological (Kuusela, Ometiță and Uc ̧an 2020) and hermeneutic (Kačerauskas 1999) motives became more frequent. The study of mental, human behavior, experiences, peculiarities of information perception, spirituality and morality required qualitatively new approaches that would take into account the subjective component, which was neglected in the process of constructing formal ideal languages. The process of empirical research also seemed completely different, because the inner world of human, as well as historico-cultural or socio-political reality were difficult to analyze in such a way as to obtain holistic and sound theoretical results without relativistic or anti-realistic ideas, which became an integral part of formal analysis. An objective assessment of the real foundations of the world and the place of human in it, the correlation between facts and values, logical and ethical, subjective and objective, required a change in the terminology.

It will not be difficult to notice how the way of representing information in Austrian thinker's later works has changed in comparison with his earlier works. If in "Tractatus Logico-Philosophicus" Wittgenstein was for the most part aphoristic, laconic, and categorical in his statements, in "Philosophical Investigations" his reasoning became more expansive and less categorical; the thinker did not express ready-made solutions as much as he carried out research, the results of which were completely unknown to him. His later works lost their pronounced systemicity and integrity, but became more critical. As a result, various elements of the narrative became clearer in his texts. Wittgenstein constructed a plot (which sometimes took the form of an imaginary situation or a thought experiment), presented ideas in a characteristic narrative style, and gave simple examples that would be understandable to every reader. The thinker did not define his thoughts as absolute truth, although he suggested such a style of investigations that set certain cognitive standards (based on language analysis). Failure to comply with these standards would mean going beyond analytic philosophizing.

There are two types of narratives used by Wittgenstein: micro-narrative and macro-narrative. Micro-narratives are usually limited to one paragraph. For example, in § 1, Wittgenstein described a language-game with a very primitive plot: someone was sent to the shopkeeper with a card that read

«five red apples» (Wittgenstein 1968, § 1). In response, the shopkeeper gave him five red apples. The peculiarity of the situation is that the reader should have question: how does the shopkeeper understand the meaning of the words «apple», «red» and «five»?

According to a similar scheme (the speech act of one person—the reaction of another person) is the story of the builder A and his assistant B (Wittgenstein 1968, § 2). The language they use contains only 4 words: «block», «pillar», «slab» and «beam». Everything that happens in this situation comes down to the fact that A exclaims one of these words, for example, «Slab!», and B in return brings the appropriate thing (in our case, the slab). B learned to bring this thing in response to the appropriate exclamation of A. That's the whole micro-narrative: again, a minimum of people, dialogues, actions.

In this way, Wittgenstein built step by step certain stories that were needed only to become the object of analysis. Thomas D. Eisele calls Wittgenstein's characteristic way of presenting thought "instructive narratives" (Eisele, 1990, p. 77) and argues that "a central part of Wittgenstein's teaching is his use of stories, or narrative" (Eisele, 1990, p. 78). However, to my mind, it should be noted that in the case of Wittgenstein's analysis of a particular situation, the analysis itself is not a storytelling. Stories are an element of analysis. Each of them can be complicated. For example, you can increase the number of words in dialogues: add numerals or adjectives and so on. However, even if we reach the complex grammar of natural language, in fact, within the corresponding language-game (according to the scheme *request—reaction*) nothing will change. Wittgenstein pointed out that a change in the actions of communicants will take place depending on the intonation with which a word is uttered, depending on who says it and the context in which it is said. If this context is clear, the action will be performed correctly, otherwise the development of events will not take place in the way that one of the communicators expects. The more events in the activities of communicators will be described, the more the narrative about them will acquire integrity.

The examples given by Wittgenstein are only isolated situations: they can be considered in themselves (as micro-narratives), and they can be inscribed in a more complex plot line (macro-narrative) that will contain new

meanings. These meanings will make it possible to understand why a certain story was actually told. In broader contexts, each story acquires a new practical purpose. In Wittgenstein, such a macro-narrative is determined by the critique of the Augustinian theory of meaning: particular stories (about buying apples and the builder and his assistant) are only a means to critique this theory. Gradually, a certain story is supplemented by the author's current reflections on a particular aspect of the philosophical question. The opinion expressed in one paragraph often develops in the following, while the author highlights a new aspect of the research question (for example, draws attention to the meaning of a word or phrase, models a new situation or considers a new context of the research problem). In this way, micro-narratives grow into a macro-narrative that contains a description and analysis of more fundamental philosophical issues. For instance, there are several of them in the "Philosophical Investigations": in addition to the critique of the Augustinian theory of meaning, it is also the rule-following, private language, linguistic means of expressing mental states, seeing aspects, and so on. To investigate them, Wittgenstein criticized the language. This approach allows not only to identify shortcomings in the reasoning of other thinkers (St. Augustine, Kant, William James, etc.), but also to point to the grain of truth in their works. However it should be noted that Wittgenstein paid very little attention to the historico-philosophical component of the question: he constructed a research narrative in such a way so as by appealing to the ideas of others to confirm the correctness of his own position.

For Wittgenstein, narrative is not a self-sufficient and self-valuable phenomenon. What is important is not the form of information, but its content. As an analytic thinker, he minimized narrative characteristics such as symbolism, polysemy, and subjectivity. It is important for him to pay attention first of all to the reality asserted in the proposition, the linguistic means used for communication and the context of the narrative, which became not a goal but only a means of constituting the individual through language. In general, Wittgenstein researched and expressed the narrative in terms of its historical, literary and scientific characteristics. This distinctive interpretation of analytic narratives differs from that proposed by Robert H. Bates, Avner Greif, Margaret Levi, Jean-Laurent Rosenthal and Barry

R. Weingast (2020), because it allows us to analyze each of them separately, rather than combine into one.

The reality of the historical narrative is determined by what happened in the past. This is the basic difference that distinguishes this narrative from the literary one. The latter does not necessarily have a reference to real events. However, artistic means (by which it is possible to create certain images and express the author's attitude to the sense of propositions) sometimes become the most effective means of formulating an opinion by the narrator. It is no coincidence, therefore, that these means are characteristic of Wittgenstein's later philosophy, in which the narrative unfolds directly in the text itself. It is noteworthy that the end of the narrative is often not known to the author himself, because the study is not yet complete. Although it is clear that his narrative does not go beyond the rational paradigm of philosophizing, because such a way out will only distance the thinker from the truth, but will not make him closer to it.

3. Narrative: from the historical past to the actual reality

Despite the fact that Wittgenstein was mostly interested in the present, the events that preceded it, of course, also had some significance for him. Study of them, based on the analysis of historical narratives, makes it possible to better understand the specifics of the state of affairs that is characteristic of today's world. Wittgenstein, following the methodology of neorealism, considered the historical process as unambiguous, although he understood that interpretations of any event are far from unambiguity. He pointed out that the confidence in the truth of certain propositions about past events is not the same. We can easily check them by referring to a specific source, which may be our memory, perceptions, memories of other people, certain records, and so on. Historical facts are scientific because we know how they can be refuted. In this regard, Wittgenstein remarked: "I am e. g. quite sure of the date of a battle, but if I should find a different date in a recognized work of history, I should alter my opinion…" (Wittgenstein 1969, § 66). In other words, the scholar will always be limited by the information that has been captured in some way. Of course, since the scientist independently determines the source reliability, it is possible that an

error may occur. However, they have no other tools to study the original sources. All that remains for the researcher, as Arthur Danto later similarly reasoned, is "analysis of historical thought and language, presented as a systematic network of arguments and clarifications, the conclusions of which compose a descriptive metaphysic of historical existence" (Danto 1965, VII).

As we can see, under the influence of new facts, according to Wittgenstein, the historian offers a new interpretation of history. In the process of creating a narrative, they are somewhat like a writer who can re-evaluate the meanings of certain events and begin in a completely different way to explain the causes of previous events. However, everything depends not on their imagination, but on the facts. This is one of the fundamental differences between narrative in history and narrative in literature. Having received new facts, the historian must in some way combine them with existing ones, that is, offer a holistic picture of what happened. Under such conditions—when when some facts conform with others—a certain subjectivity may arise, because not all causal connections between the facts can be established with certainty. Therefore, sometimes for the sake of narrative integrity it is necessary to choose the most probable of the hypotheses to explain a certain sequence of events. In this case, the degree of subjectivism will be the smaller, the less event evaluation will differ from what took place in reality.

Wittgenstein realized that in principle we can doubt the truth of certain historical evidence, and in many cases all we can do is believe that it is true. Similarly, we believe the physical, geographical, and astronomical facts taken from books. Doubt about these facts seem pointless, because there are so many things to do with them (Wittgenstein 1969, § 312). For the construction of a historical narrative, as for any other human knowledge, it is important to be convinced of the truth of certain propositions that would not be in doubt. These are peculiar propositions-hinges—all other knowledge is based on them. Explaining this view, Wittgenstein wrote: "It strikes me as if someone who doubts the existence of the earth at that time is impugning the nature of all historical evidence" (Wittgenstein 1969, § 188). We cannot say that historical evidence is definitely correct, but it is not appropriate to deny the truth of those facts without which history itself would not be possible. In this regard, the Austrian philosopher

clarified: "What we call historical evidence points to the existence of the earth a long time before my birth;—the opposite hypothesis has nothing on its side" (Wittgenstein 1969, § 188). The difficulty is to draw the line between what is certain and what can be doubted. It seems to be the optimal way to determine such a distinction in a pragmatic way, i. e. taking into account the practical consequences that a certain interpretation could lead to.

In order to clarify the line between accurate and inaccurate knowledge, it is not superfluous to pay attention to such things as linguistic means of representation of thinking, logical component of methodology, as well as to compare languages of history, ontology and epistemology. But in any case, the researcher will analyze the propositions about the actual world. Therefore, relativism as a methodological principle of the study of history will be inappropriate. On the other hand, according to Wittgenstein, it could hardly be said that the historical narrative appeals to some substantiality. In this case, the narrative should be subject to certain laws. However, Wittgenstein is known to be very skeptical about the possibility of predicting future events based on knowledge of certain laws: "A necessity for one thing to happen because another has happened does not exist" (Wittgenstein 1922, § 6.37). We can only make certain assumptions and formulate certain laws that are consistent with our experience. Having discovered new facts, these laws can be revised, because they have not so much logical as psychological (obtained by induction) basis. Therefore, it is inappropriate to find certain analogies between events of different epochs or cultures in order to predict, because the context of each of them will be unique. In addition, we cannot know the opinions of other people who influence the course of history.

However, any historical assessments expressed in narratives must be based on certain ontology. Of cource, this does not justify the truth of the substantive philosophy of history. On the contrary, the belief that history is a certain substance leads to the recognition of certain laws that can be obtained deductively, and therefore, in this case it will be possible to extend the experience of the past to the present and future. It follows that events could be described before they occur. The fallacy of this view is due to the fact that the opinions of other people who influence the course

of history are unknown. In this regard, any predictions cannot be considered reliable.

When describing a certain event (past or present), it is important for the scientist to leave everything as it is, and not to try to identify certain patterns by manipulating the facts. Another thing is that the process of description can take place with the help of different types of language-games (such as eyewitness accounts, statistical analysis, logical conclusions, etc.), which will be about people, events, processes, and so on. In any case, when the propositions describe a specific historical reality, it is doubtful that they can be considered in themselves. These propositions form a narrative that will be intertwined with other narratives in different ways. For example, they may contain concepts that are derived from the generalization of various facts into a whole (the same concept of "history" is a kind of generalization, or, for example, the concept of a particular historical period, phenomenon, process). The same people or events may be present in other narratives. Under such conditions, it is important that different narratives do not contradict each other, because historical reality is one.

4. Literary component of narrative description and its historico-cultural context

The description of any event, phenomenon or process is often not devoid of various literary techniques, such as metaphors, which give greater expressiveness to thoughts. It can hardly be argued that developing a pragmatico-analytic approach in epistemology, Wittgenstein would criticize metaphors by means of linguistic analysis, just as it could be done by means of logical analysis. Moreover, in the late period of his work, Wittgenstein himself often used metaphors. At first glance, usage of metaphors was clearly inconsistent with the methodological requirement of logical accuracy in analytic philosophy. For example, in his philosophy there are metaphors such as a beetle in a box (to denote a situation where we do not know what meanings other people attribute to words) (Wittgenstein 1968, § 293), or a fly looking for a way out of a fly-bottle (similarly, scientists on the basis of language analysis seek a way to solve philosophical problems) (Wittgenstein

1968, § 309). It is clear that the definition of philosophical issues or aim in philosophy in this way does not come close to a true understanding of the essence of things. They have a different purpose (pedagogical and pragmatic) and are designed to excite the imagination of the reader in order to stimulate their reasoning about the nature of philosophy and its objectives (Burbules 2017; Synytsia 2020). In addition, Wittgenstein's philosophical reflections are full of numerous images of learning, with which he illustrated various theoretical aspects of the issues under study—for example, the head of a duck-rabbit or a double cross (to describe a situation whose meaning may vary depending on the aspect-seeing), a schematic triangle (to demonstrate the fact that, if you do not give syntactic means clearly defined semantics, they can take on any meaning and for each of them there is a justification) (Wittgenstein 1968, ch. XI). Wittgenstein's usage of various literary techniques stemmed from the need to analyze ordinary language, as it became clear that meanings function within the language itself, rather than being derived empirically from a study of the current state of affairs.

From a pragmatic point of view, it is important that literary means make it possible to better place worldview accents and enhance the expressive effect of thought. At the same time, they inform us about the historico-cultural aspects of the narrator's life. Based on the analysis of the text, we learn about the author's intentions, his/her beliefs, interests, character, and the state of contemporary science and culture, the popularity of certain ideas, concepts or thinkers. For example, analyzing his later works, it will not be difficult to notice how he was interested in psychology (albeit in anti-psychological interpretations), as well as the actual importance of the study of consciousness as a precondition for language (see: Wittgenstein, 1968, ch. XII). He understood that the reasons for the formation of concepts can be sought not only on the basis of language analysis, but also in another way (by analyzing what is in their nature), although he added that "we are not doing natural science" (Wittgenstein 1968, ch. XII). That is, on the one hand, Wittgenstein adhered to the scientific requirements for philosophizing, and on the other hand, given the unsatisfactory state of contemporary neuroscience, he realized that some approaches to the study of consciousness are not yet available to scientists. In view of this, the research narrative of

the analytic philosopher contained only those arguments that were substantiated and confirmed by the science of the time.

It can be argued that in various narratives—not only in philosophical works, but also in religious, legal, literary ones, the narrator expresses certain linguistic and psychological structures, verbalizing their own intentions, thoughts and experiences. Narrative does not exist in itself, but is an element of a system of discursive contexts. Therefore, the narrative—individual or collective—manifests the discourse as a narrator's way of thinking and acting. The method of analysis in itself does not always make it possible to see a certain integrity of the text, the affinity of some considerations with others. However, by making research more pragmatic, Wittgenstein significantly expands the possibilities of analytic methodology, which also begins to perform synthetic, constructive tasks and through the study of various forms of narrative deepens our understanding of cultural and historical discourses of certain times.

Forms of narrative (a kind of grammar of language in Wittgenstein's interpretation) represent our ways of perception and determine certain ontology. Narrative becomes a set of prescriptive norms that reconcile and fit individual statements into a much broader historico-cultural context. Different people fill this context with their own local meanings, so they may perceive and express the same narrative differently. This is well illustrated by the example of primitive communities. Investigating the ways in which information is interpreted in these communities and its symbolic forms of expression, Wittgenstein, as Gunter Gebauer notes, clearly realized that "signs derive their meaning from shared human practices" (Gebauer 2017, 75). Thus, by examining these practices, it would be possible to better understand the nature of meanings and in general the narratives in which they are represented (Wittgenstein 1993). Wittgenstein did not evaluate the practices of primitive people as erroneous or unwise. It is unlikely that people could not notice that natural phenomena occur without their magical actions. Their actions and narratives are always full of symbolism. It is not good to explain ritual actions from the point of view of science, because they appeal to other worldview. Such practices are important because they are significant to individual communities. These practices are original, they are not developed in the same way as science (progressively). Religious and

scientific ontological hypotheses, which state the order and homogeneity of phenomena in reality, coincide. Therefore, according to the logic of Wittgenstein, one worldview is not worse than another. Various magical practices are a specific form of life that allows people to organize their activities. They capture human cognitive experience in specific narratives and can be interpreted as socio-cultural practices that do not require scientifico-theoretical explanations and generalizations. Therefore, it is not advisable to evaluate them as true or false. Human activity in primitive society is one of the language-games, a form of life that is full of its own meanings and intentions. The analysis of early human practices from the standpoint of modern European civilization will be too biased. We have completely different requirements for the ways of expressing opinions and their justification.

In general, the way in which the narrative is expressed affects the perception of listeners. By mastering various language-games, a person learns to express their own thoughts in narratives, the structure of which is flexible and open to change. Narratives become certain cognitive models of reality, which is formed in certain contexts and makes it possible to manifest human rationality as a characteristic feature of public (collective) discourses. For its part, collective discourse becomes manifested as a language-game. The very process of communication, which is the primary reality for the human person, is always social in nature. It is manifested not only through verbal means, but also non-verbally. However, this interpretation of the narrative through the prism of discourse seems to be ambiguous. Narratives have authors, while discourse is impersonal. The closer the concepts of discourse and narrative converge at the substantive level, the more it will become impossible to apply a formal analytic methodology for their study. Moreover, in the case of some narratives that are based on historical facts, their partial verification will be possible, and in the case of other, fictional narratives, such a procedure will be impossible.

From the form of narrative presentation of information depends on what meanings the listener will extract from it. The very morphology of the story and the vector of unfolding of the plot lines are constituted by its end, which is known only to the narrator. In Wittgenstein, the narrative is very often with an open end and not completely clear plot. The form of

presentation of the opinion he chose does not determine the content (it is determined by the facts) of the text, although it forms the reader's view of things, sets the direction of their interpretation of proposition. Sometimes such an interpretation is based on the author's explanation, which develops the idea mostly deductively and associatively, but not intuitively or arbitrarily. In reality, events develop independently, even if no one influences them.

The narrative developed by Wittgenstein takes the form of research. According to his theoretico-methodological approach, the best way to start a narrative is to ask questions. In "Philosophical Invesstigations" Wittgenstein asked questions more than a thousand times. Another thing is that he left most of them unanswered. However, even if he offered an answer, it did not always seem quite right to him. Thus, formulating the question, at first glance it may seem that Wittgenstein himself did not know where this or that line of reasoning would lead him. This manner of constructing a narrative can be explained by the influence of his school practices. In formulating his opinion, Wittgenstein seems to have a dialogue with the students. Interestingly, sometimes he did just that—dictated philosophy (Gibson, and O'Mahony eds. 2020). Even the examples that Wittgenstein gives to explain his opinion are very often related to school (e. g., series of numbers, grammar exercises, graphical images, and basic information on geography, history, literature, etc.). It is important for him not just to find something himself, but to do everything to encourage students or readers to search independently. He said about his teaching method: "In teaching you philosophy I'm like a guide showing you how to find your way…" (Gasking and Jackson 1967, 52). Thus, the addressees of his messages were given an active role in the creation of the narrative, which was by no means limited to the form of presentation. Wittgenstein set only those rational limits within which readers should move on their own. In this regard, he noted: "Anything your reader can do, leave to him" (Wittgenstein 2006, 77e). The narrative can be fully understood only when, figuratively speaking, to act as its co-creator. Consequently, finding all the meanings that are represented in the narrative can only be the result of a detailed and meticulous analysis, that is, when you begin to understand the very idea of the speaker.

5. Narrative in science and narrative of science

Since the narrative is a relatively independent phenomenon, the question of its scientific research arises. In particular, Wittgenstein drew attention to the fact that the narrative itself has an impact on the human mind. In § 524 of the "Philosophical Investigations", he noted: "Don't take it as a matter of course, but as a remarkable fact, that pictures and fictitious narratives give us pleasure, occupy our minds" (Wittgenstein 1968, § 524). This means that narratives are one of those things that can amaze. They make you think and evoke emotions. Narratives affect human behavior. Therefore, in order to influence a person, it is important to choose the right communication strategy and build the narrative in such a way that it evokes certain mental states, especially thoughts or emotions. This already means that philosophy as a set of narratives cannot leave everything as it is. And what is important, these narratives are not only born of surprise, but are also capable of surprising (and by no means their formal presentation, but their imagery). For this purpose philosopher does not need to create imaginary worlds—he or she just needs to focus on the study of the actual world.

Thus, investigating the narrative, the analytic thinker begins to more thoroughly understand the specifics of cognitive practices. Through the prism of the study of narrative as a means of communication in society, the researcher gradually comes to the study of linguo-psychological aspects of the mind-body problem in general and the mental in particular. At the same time, it is clear that at the personal level, the mental has a socio-cultural basis—it is not appropriate to describe it exclusively as a set of neural processes in the brain. The fact is that we represent our lives by applying a certain narrative. Thus, our understanding will be determined by a narrative that will not only represent reality, but will also be a means of its social construction. However, from a psychological point of view, it is not entirely expedient to emphasize only the socio-cultural connotations of the narrative. Its more important characteristic will be the capture and expression of intentions to act in a certain way. Moreover, these intentions may be unconscious, which significantly complicates the process of their study. In order to correctly interpret different kinds of intentions, it is necessary to develop a correct methodology that would take into account that in the world

there are no stories, but certain events (cf: Bruner 2004, 691). We construct a narrative by which we describe various events in our lives. Narrative, if we interpret Wittgenstein, becomes the form through which we express our life and everything that happens in it.

Unfortunately, the correct scientific interpretation of the narrative will be complicated by the fact that our experience can be represented in the narrative in different ways, sometimes contradictory. The point is the facts that form the plot of the narrative are not always interconnected. Therefore, the meanings inspired by them will not always generate identical interpretations, and the narrative in empirical dimensions will become relative and uncertain. The fact that over time, past events will be interpreted differently does not mean that they will be falsified. With more attention to the facts, the interpretation of these events will become clearer. However, it is also possible to lose the connection between the images of memory and the impressions that caused them. Under such conditions, the narrative becomes closed to the essence of the events it describes and can be used in different cultural contexts to argue exactly those values that are characteristic of the community.

The relativity of the narrative is generated not only by the requirements for its construction, but also by the language itself, which reflects our inherent way of life. The human worldview is so unique that it does not allow us to understand other species. It is no coincidence that Wittgenstein stated the following: "If a lion could talk, we could not understand him" (Wittgenstein 1968, ch. XI). Of course, this assumption cannot be taken literally, because at the narrative level, each conversation has a certain logic (causation, connotations, reference, etc.), without which language would not be a means of communication. It is rather about the possibility of a clear understanding of all the meanings that are embedded in the language. The structure of language (regardless of the correlation between real objects) has its own ontology, defined by the way of human life. Therefore, even if we knew a language unknown to us, but the traditions followed by its speakers would be foreign to us, we would still not be able to understand them. The fact is that we would first have to adapt to their way of life, ignoring our own previous experience.

Such Wittgenstein's skepticism about the possibility of a reliable narrative reconstruction of events can be considered more broadly, in the context of the problem of language comprehension. Indeed, we can consider any expression of language as metaphors. The language of science is no exception. The basis of understanding between people who speak different languages is most likely trust, not a certain explication. Empirical experience can be expressed within various theories (conceptual schemes). This state of affairs indicates that the language of science becomes a kind of narrative, and every scientific fact, in addition to certain information about reality, also informs about certain cultural values of the intellectual community. Such pluralism, which provides grounds for competition between scientific programs as certain types of theoretico-methodological narratives in order to be recognized more precise in explaining the ontological and epistemological foundations of reality, agrees well with pragmatic methodology, because when the scientific community prefers a scientific program, it is important to consider those practical consequences which are a defining argument for substantiation of theoretical positions. If the scientific program cannot be verified empirically, we will have to appeal to all sorts of abstract (metaphysical) concepts or thought experiments. The pragmatic conclusion would be that if metaphysical knowledge is useful for scientific theory, it is impractical to neglect it while constructing the narrative of science. For its part, thought experiments can be the only means of at least partially clarifying the situation referred to in the narrative, in cases where empirical experiments are not possible.

In the light of this kind of perspective of philosophical research, the question of truth has arisen in a new way. Analyzing linguistic expressions, Wittgenstein remarked that quite often it may seem that "the feeling gave the words *truth*" (Wittgenstein 1968, § 544). In other words, the meaning of words will depend on what we experience and feel when we make certain statements. The defined boundaries of truth become somewhat blurred. As Richard Rorty noted, truth can become dependent on the human mind and on the way how we describe the world around us (Rorty 1989, 5). The dependence of truth on the narrative makes it a kind of phantom "grand narrative", which breaks down into many truths of each individual. Under such conditions, the narrative itself, rather than objective reality, sets the

meaning of truths. If this were the case, it would mean the complete dependence of truth on the language of scientific theory, its to some extent personification and relativity, as well as the possibility of permanent creation of new meanings in new contexts and, in general, pluralism of interpretations of intricate narrative. However, Wittgenstein did not seem to deviate from the principles of objectivist methodology, and therefore argued quite pragmatically that: "An 'inner process' stands in need of outward criteria" (Wittgenstein 1968, § 580). If certain feelings arise, they become conditioned by some external situation. The researcher's attention should be focused on the knowledge of external realities in order to get closer to understanding the nature of things, even though, by perceiving or analyzing the narrative, we can involuntarily influence its results. In any case, by clarifying the pragmatic aspects of the narrative, one can much more clearly investigate the specifics of intersubjective communication and more thoroughly understand the peculiarities of perception of reality.

6. Narrative: openness to discussion

Wittgenstein's analysis of narratives encounters a number of difficulties that need to be analyzed. In particular, it is about the correlation between language-game and narrative. It seems that not every language-game is a narrative, however, it can be reformulated into a narrative. To do this, we need to consider the language game more broadly—given the speech context. In other words, the situation of retelling a language-game becomes a narrative that has the characters, a story line, and so on. The end of such a narrative depends on the extent to which the requirements for language communication have been met. In addition to the meanings, it is also important who were the speaker and the listener, what were their intentions and beliefs, under what circumstances the communication took place—all these components are important for the development of the plot. According to Wittgenstein, language-game is «the whole, consisting of language and the actions into which it is woven» (Wittgenstein, 1968, § 7). Given the possibility to use action to influence what is really happening, language-game becomes a way of life. Moreover, with its help it is possible not only to influence, but also to describe the reality and the changes that take place

in it. In this case, the language game acquires certain elements of the narrative, which is characterized by an expressive function of influence. And where there is the influence of one on the other, there is an opportunity to develop a story line, that is, to construct the development of events. They are suggested by the narrator, but the reader still needs to make an effort to correctly interpret the proposed story.

If the story begins with a question, it is not clear whether its end may be unknown to the narrator? In the case of Wittgenstein, things are not so clear. It is clear that the text presented by him was already well thought out from the very beginning. Wittgenstein only suggested that readers find this logic themselves, and not come up with their own interpretation of the arguments. However, this applied only to the logical component of the narrative. In addition, each narrative has aesthetic and emotional components. The latter two can be interpreted differently by researchers. After all, those associations that are evoked by the narrative can differ significantly from one reader to another. Differences in interpretations of the narrative are due to life experience, beliefs and interests, as well as cultural and historical background.

In this case, Wittgenstein's analysis of the narrative cannot be considered outside the context of cultural studies (McDonald, 2001). Moreover, as Jens Brockmier and Rom Harre argue, studying its narrative, its discursive embeddedness cannot be neglected (Brockmier & Harre, 1997, p. 264). Such embeddedness is manifested through the prism of cultural and historical landmarks for the development of thought. However, how can we talk about the discursive embeddedness of Wittgenstein's narratives? He sought to build a narrative that would be as general and understandable as possible to a wide range of readers. In this way, the cultural and historical preconditions of the story seem to be eliminated. The way of presenting an opinion becomes close to scientific, but scientific texts are still not devoid of narrative structures. It becomes unclear where the line is between the description of reality and the process of its construction.

7. Conclusions

In general, it can be argued that Wittgenstein in the late period increasingly paid attention to the pragmatic aspects of expression, in particular, creating a narrative. The micro- and macro-narratives present in his works are important elements of the analysis of language from the pragmatics standpoint. Wittgenstein focused the reader's attention on a particular story, but did not always draw clear conclusions from it. Stories became the beginning of philosophizing. In general, in his philosophical studies can be distinguished analysis of aspects of historical, literary and scientific narrative. Each of these narratives acquires its own characteristics. This approach complicates the analysis of the narrative, but makes it more thorough, holistic and logically thought out, that is, one that would correspond to the current state of affairs in reality. It is no coincidence that Wittgenstein avoided relativization and subjectivity in interpreting the events of the past, although he understood that historical statements could change under the influence of new facts. Wittgenstein was well aware that the way people interpreted an events would be influenced by historico-cultural context and human experiences. Despite the fact that narratives have their own logic of presentation, it is not advisable to look for any patterns in the development of history, as the future is uncertain. From a pragmatic point of view, narratives cannot be analyzed as true or false because they are not devoid of metaphors. However, this does not mean that the author has the right to interpret events at his own discretion—it is important to describe them as clearly as possible. Metaphors are needed to make the ideas clearer, not less true. It is important for Wittgenstein to be as clear as possible, not to confuse the reader. Imagery is important for this, but first of all the depth of thought must be a decisive argument. From a pragmatico-analytic perspective, in the process of researching the narrative, it is important to capture the meanings that are hidden in speech, rather than invent new ones. Also, capturing the intentions and beliefs of the author, the researcher must take into account that they are inseparable from the discourse of a certain time. Their narratives must be consistent with standards of analytic precision. Similarly, the narrative itself can be studied from the standpoint of science. At the same time, it must be realized that a reliable

reproduction of all the objective realities which determine the narrative is hardly possible. However, as Wittgenstein has shown, a pragmatic approach to analysis allows us to deepen our understanding of the nature of reality and human existence, the knowledge of which we express in narratives, and thus to influence ourselves through the process of narration.

References

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 2020. *Analytic Narratives*. Princeton: Princeton University Press.

Boncompagni, Anna. 2016. *Wittgenstein and Pragmatism: On Certainty in the Light of Peirce and James*. London: Palgrave Macmillan.

Bruner, Jerome. 2004. "Life as Narrative." *Social Research* 71 (3): 691–710.

Burbules, Nicholas C. 2017. "Wittgenstein's metaphors and his pedagogical philosophy." In *A Companion to Wittgenstein on Education: Pedagogical Investigations*, edited by Michael A. Peters, and Jeff Stickney, 123–33. Singapore: Springer. https://doi.org/10.1007/978-981-10-3136-6_8

Danto, Arthur. 1965. *Analytic Philosophy of History*. New York: Cambridge University Press.

Eisele, Thomas D. 1990. Wittgenstein's Instructive Narratives: Leaving the Lessons Latent. *Journal of Legal Education* 40 (1/2): 77–95.

Garrison, Jim. 2017. "Wittgenstein and Classical Pragmatism." In *A Companion to Wittgenstein on Education: Pedagogical Investigations*, edited by Michael A. Peters, and Jeff Stickney, 321–38. Singapore: Springer. https://doi.org/10.1007/978-981-10-3136-6_21

Gasking, Douglas Aidan Trist, and Alan Cameron Jackson. 1967. "Wittgenstein as a Teacher". In *Ludwig Wittgenstein: The Man and His Philosophy*, edited by K. T. Fann, 49–55. New York: Dell Publishing Co.

Gebauer, Gunter. 2017. The Turn to Anthropology. In *Wittgenstein's Anthropological Philosophy. History of Analytic Philosophy*, 75–107. Cham: Palgrave Macmillan https://doi.org/10.1007/978-3-319-56151-6_4

Gibson, Arthur, and Niamh Ann O'Mahony, eds. 2020. *Ludwig Wittgenstein: Dictating Philosophy: To Francis Skinner—The Wittgenstein-Skinner Manuscripts*. Cham: Springer Nature.

Kačerauskas, Tomas. 1999. "Is L. Wittgenstein Analytic or Hermeneutic Philosopher?" *Problemos* 560: 71–84. https://doi.org/10.15388/Problemos.1999.56.6863

Kuusela, Oskari, Mihai Ometiță, and Timur Uc¸an. 2020. *Wittgenstein and phenomenology*. Abingdon, Oxon; New York, NY: Routledge.

McDonald, Henry. 2001. Wittgenstein, Narrative Theory and Cultural Studies. *Telos* 121: 11–53.

Rorty, Richard. 1989. *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.

Synytsia, Andrii. 2020. "Pragmatic Aspect of Wittgenstein's Philosophy of Education: A Modern Interpretation." *Journal of Education Culture and Society* 11 (2): 39–51. https://doi.org/10.15503/jecs2020.2.39.51

Unger, Peter. 2014. *Empty Ideas: A Critique of Analytic Philosophy*. New York: Oxford University Press.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. Translated by Charles Kay Ogden. London: Kegan Paul, Trench, Trubner & Co.

Wittgenstein, Ludwig. 1968. *Philosophical Investigations*. Translated by Gertrude Elizabeth Margaret Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig. 1969. *On Certainty*. Translated by Denis Paul, and Gertrude Elizabeth Margaret Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig. 1993. Remarks on Frazer's Golden Bough. In *Ludwig Wittgenstein. Philosophical Occasions, 1912–1951*, edited by James Klagge, and Alfred Nordmann, 118–155. Indianapolis: Hackett Pub Co Inc. (in German and English).

Wittgenstein, Ludwig. 2006. *Culture and Value*. Translated by Peter Winch. Chicago: Chicago University Press.

Organon F 29 (3) 2022: 348–375
https://doi.org/10.31577/orgf.2022.29303

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

On the Appropriateness of Holding Morally Accountable

Haozhan Sun* – Ivo Dragoun**

Received: 15 October 2021 / Revised: 19 July 2022 / Accepted: 19 August 2022

Abstract: Our everyday social interactions involve holding others morally accountable for their wrongdoings. Sometimes such holdings might be inappropriate. For instance, it feels inappropriate if the person holding another morally accountable is in some relevant sense morally compromised. Thus a thief chastising a thief would strike us as somewhat odd. We might, when witnessing such behaviour, want to remind the chastiser that he is not in a position to reprimand others. But what if none of us are ever in such a position? In this paper we will argue that all men are irremediably morally compromised, and conclude that, ultimately, it is never appropriate for us to hold others morally accountable.

Keywords: Blame; holding accountable; moral standing; the Principle of Plenitude.

* Zhejiang University of Science and Technology
https://orcid.org/0000-0002-4457-7595
Zhejiang University of Science and Technology, 318 Liuhe Rd, Xihu, Hangzhou, Zhejiang, China
hs370@sussex.ac.uk

** Sussex University
https://orcid.org/0000-0003-4189-8657
University of Sussex, School of Media, Arts and Humanities Sussex House, Falmer Brighton, BN1 9RH United Kingdom
id41@sussex.ac.uk

1. Introduction

In our social interactions we often hold others morally accountable for their wrongdoings. It is generally recognized that sometimes such a holding might be inappropriate. An obvious case of inappropriately holding another morally accountable is when the holder misidentifies the agent behind the relevant wrongdoing and targets the wrong person. Similarly inappropriate is to hold someone morally accountable who had no way of knowing his actions would result in a wrongdoing. The case of an inappropriate holding of another morally accountable that we shall discuss in this paper is the one in which the holder himself is morally compromised, and as such is not in a position to criticize others for their wrongdoings. We will argue that, in fact, none of us are in the position to hold others morally accountable exactly because we are all essentially morally compromised. This argument crucially relies on the statistical interpretation of modality and a thought experiment designed to show that one's *future* wrongdoing can affect one's *present* moral status.

2. Blame, holding accountable, and the moral standing

There has recently been increased interest in the topic of *blame* among moral theorists.[1] The issues discussed in connection with *blame* are the nature of blame, its function and the conditions for the appropriateness of blaming.

Regarding the appropriateness of blaming others for their wrongdoing, there seems to be a consensus about what general facts must be taken into consideration when assessing the appropriateness. The facts relevant for the appropriateness assessment can be broadly sorted into three groups: (a) facts about the blamer, (b) facts about the blaming interaction, and (c) facts about the person being blamed. In what follows, we will make use of a concept that plays a crucial role in discussions of the first group, that is,

[1] For an excellent collection of essays on blame see (Coates and Tognazzini eds. 2013).

in discussions of facts about the blamer. The concept we have in mind here is the concept of *moral standing*.

There are plenty of real and imaginary scenarios in which even if the agent was blameworthy, and even if all procedural norms were followed, it would be morally inappropriate for some people to blame. For blaming to be appropriate, not only the target person has to be *blameworthy* but the blamer himself has to be what Friedman (2013, 272) calls *blamer-worthy.* For instance, a serial thief blaming another for a petty theft would strike us as such a case of morally inappropriate blaming exactly because the serial thief is clearly not *blamer-worthy.* 'Who are *you* to criticize another for *that*?', we would want to interject if we witnessed an instance of such a morally inappropriate blaming. What gets brought into focus and questioned in such an interjection is the *moral standing* of the blamer. The background intuition here is that if the moral standing of a person is in some relevant sense compromised then it would be morally inappropriate for that person to blame another.

Uncontroversially, holding others accountable involves blaming. Thus, to the extent to which the appropriateness of blaming depends on the moral standing of the blamer, holding someone accountable will also depend—with regard to its appropriateness—on the moral standing of the holder. If the moral standing of a holder has been compromised, then the holding is (morally) inappropriate or wrong. A husband involved in a long lasting extra marital affair attempting to hold his wife accountable for a one-night stand with her ex would strike us as an instance of a morally inappropriate holding accountable because of the compromised moral standing of the husband.

3. Moral standing of human agents as essentially compromised

The moral appropriateness or rightness of holding another accountable depends crucially on the moral standing of the holder. Now, what if the moral standing is *always* compromised? What if a fundamentally compromised moral standing is in some sense *essential* to being human? If this was

the case, surely, we would have to conclude that our practice of holding others accountable is morally, and irremediably, wrong. There are reasons to believe that our moral standing is fundamentally compromised in some such way. Below, we shall formulate an argument to this effect. The argument will employ (a version of) a thesis known as The Principle of Plenitude and a thought experiment that allows its applicability in the context of evaluating the moral standing of all men. Let us first give a rough outline of the argument so that the logical role and the mutual relation of its individual steps is clear before they are later discussed in detail.

An outline of the argument:

1. The Principle of Plenitude (PP): For some states of affairs s, if s is possible then there is a time at which s obtains.

2. A state of affairs that has obtained is a possible state of affairs.

3. Plausibly, if a state of affairs involving a member m of a kind k is a possible state of affairs then a relevantly similar state of affairs involving any other member of the kind k is possible too.

4. Some men did or have done things that (have) corrupted their moral standing.

5. It is possible for any man to have their moral standing corrupted. [from 2, 3, 4]

6. [*A thought experiment designed to show that*] One's (present) moral standing becomes corrupted (not only by the past and present wrongdoings but even) by one's future wrongdoings.

7. [*For reasons discussed below*] PP can be applied to (5), that is: the status of one's moral standing belongs to the states of affairs that PP applies to.

8. For all men there is a time at which they do things that corrupt their moral standing. [from 1, 5, 7]

Therefore

9. The moral standing of all men is essentially corrupted. [from 6, 8]

The individual steps of the argument will be discussed in dedicated subsections below.

3.1 The Principle of Plenitude

The Principle of Plenitude as we will understand it here is grounded in what is sometimes called the 'statistical interpretation of modality'.[2] The statistical interpretation (or model) of modality can be spelled out roughly in the following way: what is necessary is always actual, what is impossible is never actual and what is possible is at least sometimes actual. The Principle of Plenitude is then a thesis about a certain kind of relation between *possibility* and *actuality*. A good first approximation of the principle is given by Hintikka (1981, 58):

> [*A*]*ll genuine possibilities*, or at least all possibilities of some central and important kind, *are actualized in time*. Any such possibility thus has been, is, or will be realized; it cannot remain unrealized through an infinite stretch of time; in a sense, everything possible will happen in the long run.

A version of the principle (and/or the modal intuition behind it) has been endorsed by philosophers such as Plato, Aristotle, Epicurus, Augustin of Hippo, St Anselm, Thomas Aquinas, Giordano Bruno, Spinoza, Hobbes, Leibniz, Kant and Russell.[3] Kant, for instance, seems to endorse the principle in the following passage:

---

[2] The term 'statistical interpretation of modality' was introduced into the modern discussion of modality by Becker (1952).

[3] All the names listed above, apart from that of Aristotle, Hobbes and Russell, are the names given by Lovejoy (1936). Lovejoy was not a philosopher and the list is somewhat controversial. Thus, for instance, Hintikka suggests that it would not be correct to take Plato as an adherent of the principle. Regarding Kant, Hintikka argues that Kant endorses the principle only in his pre-critical writings. (We tend to disagree with Hintikka on this point as there is, in our view, a passage in The Critique where Kant seems to say something which is very close to the Principle. We quote the passage below). And as for Leibniz, although there is a version of the principle that Leibniz embraced, it is a version that is rather different from the one given above. In fact, Leibniz explicitly rejected the version that we shall work with here. For details on these points see (Hintikka, 1976).

> The Schema of possibility […] is the determination of the representation of a thing at any time whatsoever. The schema of reality is the existence at a given time. The schema of necessity is the existence of an object at all times. (Kant 1966, 125)

And Russell, in his characteristically clear and unambiguous style, asserts:

> One may call a propositional function *necessary*, when it is always true; *possible*, when it is sometimes true; *impossible*, when it is never true. (Russell 1956, 231)

And on the next page in the same paper, Russell casually equates the notion of *sometimes* with that of *possibility*:

> It will be out of this notion of *sometimes*, which is the same as the notion of *possible*, that we get the notion of existence. (Russell 1956, 231-32)

Obviously, the first approximation of the principle given above allows for different versions depending on what kind of possibility one has in mind. Possible states of affairs? Possible kinds? Possible particulars? The plausibility of the principle might depend on the kind of possibility that it is applied to. Barnes (1977, 184), for instance, thinks that the principle does not apply to states of affairs involving perishable particulars because out of the numerous possible ways that a perishable particular can perish only one can happen: once the particular has perished, all the remaining (and previously) possible ways of perishing become impossible regarding that particular because the particular is no more. Later in this chapter (when discussing step 7) we will address this issue as we need the principle to apply to individual humans, that is, to perishable particulars. At this point, however, we have a bigger problem to deal with. The problem is that in contemporary analytic literature, the statistical understanding of modality has been fully replaced by interpreting modality in terms of what is generally known as possible world semantics. And, modality interpreted in terms of possible world semantics does not involve any temporal references. Consequently, there is virtually no discussion of the principle of plenitude to be found in contemporary analytic philosophers' writings on

modality.[4] Thus, not only are we unable to simply appeal to the principle and proceed to discussing the following steps of the argument, we must also assume that the principle will be perceived by our readers as weird at best and totally implausible at worst. A way forward at this point is to provide an argument for the principle. One such argument can be extracted from (Barnes, 1977).

Barnes gives his argument (or what could be reconstructed as his argument) for the principle in the context of discussing Hintikka's interpretation of some of the aspects of Aristotle's treatment of modality. In his review of Hintikka's paper, Barnes is not primarily concerned with providing an argument for the principle. What he is concerned with is complementing Hintikka's interpretation of certain passages from Aristotle in which the Stagirite seems to argue for the principle with some charitable reading of those passages. In the passages, Aristotle's argument seems rather obscure, and this is where Barnes steps in offering a sympathetic reconstruction of Aristotle's thinking behind the argument.

The argument for the principle as given by Barnes relies crucially on the assumption that, roughly, *that which always obtains is necessary*, or:

A: if *s* always obtains then *s* is necessary

The assumption is, however, rather controversial. It will be pointed out that it is easily conceivable that *s* always happens and yet *s* is contingent. Moreover, the way modality is, in (A), tied to temporality seems to be uncannily similar to the way in which we wish to tie modality to temporality in (PP). Thus, (A) could be seen as begging the question. Clearly, an argument for (A) is needed, and Barnes gives us one. Unfortunately, there seem to be several problems with the argument. The most serious one is the fact that the argument depends crucially on Aristotle's definition of possibility—a definition that is rather implausible. Aristotle thinks, roughly, that something is (can be defined as) possible if and only if nothing impossible results

[4] Even though there are *historians* of philosophical ideas and their development, who discuss the principle in an analytic way. One of them is the above-mentioned Hintikka. Another one is, for instance, Barnes, whose brilliant argument for the principle will be discussed below.

from its actualization. Barnes puts this more rigorously in the following way:

> T: s is possible =(df) there is no state s^* such that s^* is impossible and if s obtains then s^* obtains. (Barnes 1977, 185)

It is immediately obvious, we believe, that (T) does not work as a *definition* of possibility, although it perhaps works as a *necessary condition* of possibility. Now, as mentioned above, Barnes's argument for (A) turns on the truth of (T). As a *definition* though, (T) is false. This collapses the whole argument for (A) as formulated by Barnes.[5] A different argument is needed. We have thought hard about (A) and we have come to suspect that there is no good argument showing how something's necessity could be *conceptually* derived from that something always obtaining. This, however, does not mean that (A) cannot be argued for. An alternative and common way of arguing for a claim is to show that accepting its falsity is theoretically too costly.

So, what are the theoretical costs of denying (A)? One way of denying (A) is this:

A*: s always obtains and (yet) s is not necessary.

First, notice that (A*) is a claim that will be embraced by a Humean. A Humean believes that the world has no nomological structure, that is, she does not believe that there is any causal law-likeness out there in the world. She sees only contiguity, temporal priority and constant conjunctions where others see Laws of Nature being instantiated. Uncontroversially, the notions of a '*nomological* structure' and the 'Laws of Nature' entail an appeal to the modal property of necessity. Thus, a Humean's denial of the world having a nomological structure or of (being governed by) the Laws of Nature entails her denying that there is any (physical) necessity out there in the world.[6] Consequently, she will endorse (A*).

---

[5] Barnes is, of course, aware of the implausibility of (T). The task he sets himself in the review is to reconstruct what he thought was Aristotle's argument. Barnes's faithful reconstruction inherits the problematic definition of possibility from Aristotle.

[6] This is not to say that a Humean denies *conceptual* necessity. She might do so, or she might not, depending on her other philosophical commitments. We focus on

The Humean view according to which there is no nomological structure (or Laws of Nature) behind the constant conjunctions that we observe is called a *regularity theory*. A regularity theory holds that there is nothing *beyond* the regularities that somehow hold the world together: nothing that *underlies* them, nothing that *explains* them. Now, it should be noticed that a regularity theory is a rather extreme theory. It implies that *s*'s regular and exceptionless obtaining is not grounded in the necessity that is intrinsic to the Laws of Nature but due to—given an infinite time—a mindbogglingly extraordinary coincidence. Strawson points out this implication here:

> According to [regularity theories]…, the regularity of the world's behaviour is, in a clear sense, a complete and continuous fluke. It's not just that we don't know whether or not there is any reason for it in the nature of things. According to [regularity theories], there is definitely no reason for it in the nature of things. (Strawson 1985, 21)

Strawson has no patience with this view calling it 'utterly implausible' and 'absurd':

> [T]he theory is utterly implausible in asserting categorically that there is no reason in the nature of things for the regularity of the world … it is absurd to say—to insist—that there is definitely no reason in the nature of things why regularity rather than chaos … occurs from moment to moment. (Strawson 1989, 21–2)

We are in agreement with Strawson here. There is more to be noticed about the view though. Regularity theory seriously clashes with one of the foundational assumptions behind any theorizing about the world: the assumption that the world can be made sense of; that it is *explicable*. In such a world, coincidences of this magnitude cannot exist. The assumed explicability of the world commits us to assuming that the world has a nomological structure, appeals to which play crucial role in anything that counts as an explanation of the world. What is at stake here is not only the project of natural sciences to understand and explain the world, but also the prospects

the *empirical* necessity here as we are arguing towards a modal claim about humans, that is, about empirical entities.

of *philosophizing* about it. A Humean—that is, a regularity theorist—rejects any talk of nomological structures or Laws of Nature because she denies the reality of causation. Without causation, there is no real *connection* between things that would allow the transfer of necessity between them. Thus, without causation, there are no *Laws* of Nature and nothing *nomological* about the world. But one cannot philosophize without taking causation seriously. As Beebee—who is a regularity theorist—admits,

> [there is a] huge range of fruitful philosophical theories that *do* appeal to causation: we have causal theories of perception, reference, action and knowledge; functionalist theories of the mind, consequentialism; and so on, and on. (Beebee 2006, 510)

A Humean is a philosopher. Thus, endorsing the regularity theory will cost her dearly: Not only will she have to embrace a rather unattractive claim that natural sciences don't *explain* but merely *describe* the world, she will also have to ditch (on pain of being inconsistent) a 'huge range of fruitful philosophical theories', and, perhaps, even stop philosophizing altogether because, as Beebee (2006, 510) notices, 'trying to do philosophy without ever using the concept of causation is practically impossible'.

Now, it is somewhat surprising that Strawson's objections, given how damaging they are, have not elicited much response from regularity theorists. A notable exception is the response given by Beebee in the paper we have quoted from above. The paper presents accurately and fairly Strawson's objections to regularity theories and raises several important and correct points about the objections that a Strawsonian should take into account. That's not all. Beebee takes on what we think is Strawson's strongest objection—an objection that plays a key role in our argument for (A). Above, we have appealed to the extreme implausibility and unpalatability of the implication (of regularity theories) that the highly complex orderliness of the world is a result of mindbogglingly extraordinary luck. Strawson invites the reader to appreciate the absurdity of the implication through considering the following analogy:

> [Imagine that] a true randomizing device determines the colour value of each pixel on a standard 800 x 400 computer screen, running on a ten-times-a-second cycle—so that each pixel can

> take any colour value for each $1/10^{th}$ second period. On the screen it appears that there is a film showing. A woman enters a house, walks over to a stove, and puts on a kettle. Life—a world, as it were—goes on in an ordered, regular fashion, exactly as regularly as in our own world. But the image is being generated by the true randomizing device. It is pure fluke that what happens on the screen appears to tell a coherent story of a regular, ordered world, rather than filling up with—or suddenly switching to—a fizz of points of colour. (Strawson 1989, 24)

The analogy is powerful, vividly exposing the theoretical costs of endorsing a regularity theory. Beebee's response is smart. She does not attack the analogy itself - in fact, she urges a regularity theorist that she 'must […] accept that from a metaphysical point of view the analogy is a pertinent one' (Beebee 2006, 527) - she, instead, argues that the implication can be *tolerated*. Her argument in this respect is, roughly, this: True, a regularity theory comes at the cost of accepting that the highly complex orderliness of the world is due to just massive—and ongoing—luck. However, that's nothing to be much upset about because we have already learned how to tolerate 'outrageous runs of luck' (Strawson 1989, 26). Consider your own life. You are alive as a result of an extremely long series of lucky events. Think of all those things that had to happen in order for you to be born. On countless occasions, your parents might have done something that would have prevented you from having been conceived, or they might have not even met in the first place. The same goes for your grandparents on whose actions the existence of your parents—thus yours too - depends. Ultimately, your existence and everyone else's existence depends on that spectacularly lucky streak of events that resulted in Earth being a place that supports life. Now, when you start thinking about all this, how much does it really bother you? Most likely, not much at all. You do not really think there is, or must be, any 'reason' why things happened in a way that ultimately led to your coming into existence.

This is an intriguing reply even though we do not think it works. Let's have a closer look at what is going on here. Strawson formulates a thought experiment designed to expose the extreme implausibility of the claim that a purely random process can, at the same time, be a process that is highly

ordered and keeps being so for very long periods of time. In response, Beebee invites the reader to consider their own life to see that it, too, despite being ordered and coherent, is a result of a long series of lucky events. She notices that it does not seem to bother us much that luck plays such a fundamental role in our lives and concludes that we, in fact, already know how to tolerate the seemingly intolerable implication of regularity theories. Beebee's reply hinges on being presented as an *analogy* to Strawson's thought experiment. And if it is an analogy then whatever the analogy shows can be—by analogy—said to be shown about the thought experiment. In our case: if we stay unperturbed about our lives being a massive fluke (as shown by the analogy), then why be perturbed about (a) a movie with a coherent story being generated by a randomizing device, and (b) about, ultimately, the implication of regularity theories? The problem with the reply is that the little consideration that Beebee offers as an analogy to the thought experiment is an analogy only seemingly.

The intuitive force of Strawson's thought experiment depends crucially on contrasting a *true randomizing device* with the high level of orderliness and coherence of a movie. Beebee, however, does not mention any randomizing device at all. She, instead, talks about *luck*, and contrasts it with the orderliness and coherence of one's life. Presumably, a *true randomizing device* and *luck* are treated as conceptual analogues here. *Luck*, however, is a notion that is ambiguous in a way that a *true randomizing device* is not—it has both a metaphysical and an epistemological reading. A *true randomizing device*, on the other hand, has (at least in the context of the thought experiment) only a metaphysical reading. Now, it certainly feels perfectly natural to *see* your own life as a *miracle* of sorts: so many things could have gone wrong over such a long time, and if they had, you would not have been born. But they did not, and once you stop for a moment to appreciate this, you cannot but feel extraordinarily *lucky*. This feeling, however, is just due to you having an epistemological access to only a tiny fragment of the total facts that, as a whole, *produced* you. If you knew the totality of those facts you would feel about as lucky as you feel when you clap your hands and it produces a sound, that is: you would not find anything lucky about it. The huge gaps in your knowledge about the past events that gave rise to your existence give a strong impression that those relatively very few

facts you know are somehow floating in the ocean of randomness. This ocean of randomness, however, is, in fact, just your inevitable ignorance regarding the totality of events that, ultimately, produced you. If you knew the totality of the facts, the ocean would evaporate. Beebee's reply by way of offering and exploiting an analogy to Strawson's thought experiment does not work because Strawson contrasts a highly ordered and coherent state of affairs with *metaphysical* luck, while Beebee contrasts it with (what we call here) *epistemological* luck.

In a little bit more detail, the problem is this. Arguably, most people know, often perhaps in some pre-conceptual way, that the kind of *luck* they accept as being involved in their own and others' lives is something like the kind of luck we have qualified above as *epistemological* luck. Or, *at least*, they would resist understanding the luck involved in their lives as being conceptually equivalent to a *true randomizing device*. Why do we claim this? It is a safe bet to expect that the vast majority of people would find Strawson's thought experiment convincing.[7] That is, they would agree that highly ordered and coherent states of affairs lasting for long periods of time cannot emerge out of a *truly random* process. They, at the same time, accept that they are very lucky regarding their lives, which can only mean that people normally do not understand *luck* as a *truly random* process.[8] Thus Beebee cannot claim that we already know how to tolerate luck in our lives in the sense of tolerating the thought that our lives emerge out of a truly random process. She can claim so only in terms of (what we call) epistemological luck. In this sense, however, her reply misses the target.

To the best of our knowledge, Beebee's response to Strawson's thought experiment is the strongest challenge to it that can be found in philosophical

[7] Recall that Beebee herself agrees that the thought experiment 'is a pertinent one'.

[8] The following uncharitable answer is possible: The majority of people have inconsistent intuitions. Therefore, they would be convinced by Strawson's thought experiment and yet see their lives as lucky in the sense of being truly random. This reply is rather unattractive. Accepting that most people have inconsistent intuitions severely undermines any appeal to intuitions in philosophical arguments. Some of the most important arguments in philosophy rely on an appeal to (rational/conceptual) intuitions. Thus, the uncharitable reply would be far too costly.

literature. It fails nonetheless. Consequently, a Humean objection to (A) fails too.

Once the truth of (A) has been established, the rest of the argument for (PP) is relatively straightforward. Recall:

> A: if *s* always obtains then *s* is necessary
>
> The Principle of Plenitude (PP): For some states of affairs *s*, if *s* is possible then there is a time at which *s* obtains.

(PP) is then readily derivable from (A) in the following way. Suppose that *s* is a contingent state of affairs. That is, suppose that *s* is a possible, not necessary, state of affairs. Then *s*'s not obtaining is a contingent (that is, possible but not necessary) state of affairs. If that is so, then *s*'s not obtaining is not necessary. And from this it follows, by (A), that '*s*'s not obtaining does not always obtain; hence there is a time at which the non-obtaining of *s*'s not obtaining obtains; i.e. there is a time at which *s* obtains' (Barnes 1977, 185). This might feel too condensed, so let us unpack it here a little:

i. Suppose: *s* is (a) contingent (state of affairs).

ii. (i) entails that *s*'s not obtaining is contingent.

iii. (ii) entails that *s*'s not obtaining is not necessary.

iv. A: if *s* always obtains then *s* is necessary.

v. By (A): if *s*'s not obtaining is not necessary, then *s*'s not obtaining does not always obtain.

vi. If *s*'s not obtaining does not always obtain, then there is a time at which *s*'s not obtaining does not obtain.

vii. To say that 'there is a time at which *s*'s not obtaining does not obtain' is to say that 'there is a time at which *s* obtains'.

We take it that it has been proved that (for some states of affairs):[9] if *s* is (a) contingent (state of affairs) then there is a time at which *s* obtains. This

[9] A reminder: (PP) does not apply to all kinds of states of affairs. As mentioned above, (PP) will not work for states of affairs involving, for instance, perishable particulars. This deficiency will be addressed when discussing step 6 below.

is, for our purposes here, close enough to (PP), thus we shall, from now on, assume the truth of (PP).

Before we proceed to the discussion of step (2) regarding the argument behind the claim that the moral standing of all men is essentially corrupted, let us briefly address a little issue regarding the plausibility of (PP). We strongly suspect that many readers—even those that have been convinced about the validity of the argument - will find (PP) and its implications just too fantastic to swallow. As Barnes notes:

> According to (PP), elephants will tell each other human jokes, the first daffodils of autumn will appear when the leaves fall upwards to the trees, and pigeons will hunt cats through city backyards. (Barnes 1977, 184)

This, indeed, is a rather unpalatable corollary to (PP). Is it, however, unavoidable? It is not, we believe. (PP) is a thesis about (some) possible states of affairs, i.e. it tells us something interesting (and perhaps unexpected) about (some) possible states of affairs. It does not, however, come with any prior commitment to what states of affairs count as *possible*. In this respect it is entirely up to the reader to decide what states of affairs she accepts as possible. In other words, the reader can, if she wishes, avoid the above mentioned unpalatable corollary to (PP) by refusing to accept as possible states of affairs the ones in which elephants tell each other human jokes, the leaves fall upwards to the tress, and pigeons hunt cats through cities. The authors of this paper are among those far from sure that these count as possible states of affairs. A deeper point here is this. (PP) could be taken as a metaphysical definition of possibility—this is, we believe, just a matter of choice. Once accepted as a metaphysical definition of possibility, it preempts any objection that appeals to the implausibility of its implications of the kind Barnes gives because the logic of such a definition serves as a constraint on what counts as a possibility in the first place. Be that as it may, the kind of possibility we will apply (PP) to below is nowhere close to as fantastic as the possibilities conceived of by Barnes.

3.2 A state of affairs that has obtained is a possible state of affairs

There is an obvious sense in which this assertion is correct. If *s*'s not obtaining is necessary, then *s* never obtains. Hence, if *s* has obtained then *s*'s not obtaining is not necessary. And, if *s*'s not obtaining is not necessary then *s* is possible. This little argument appeals to a straightforward logical relation between the concepts of 'something never obtaining', 'something being necessary' and 'something being possible'. The verbal phrase 'has obtained' implies - due to the usage of the action verb 'obtain' - that there was a time when the state of affairs did not exist. This rules out the possibility that the state of affairs is a necessary state of affairs.[10] So far so good. However, there is a sense in which the assertion is false. It seems clear that a temporally determinate proposition referring to a past state of affairs is, if true, necessarily true because it is metaphysically impossible to change that state of affairs. Thus the proposition 'Donald Trump lost the presidential election in 2020' is necessarily true because it is true and the state of affairs it refers to cannot be changed. At the same time, the proposition 'Donald Trump lost the presidential election in 2020' can be understood as referring to a state of affairs that has obtained. Clearly, Donald Trump's losing the presidential election in 2020 has (as a state of affairs) obtained. But we already know about this state of affairs to be necessarily true. Thus we have here an example of a state of affairs that (had not always been the case and) *has obtained* and is a *necessary* state of affairs (as it cannot be changed). The two senses in which the assertion can be understood suggest that there are at least two kinds of necessity that can be associated with a state of affairs. A state of affairs that has obtained is necessary (a) in the sense that its past obtaining is fixed and cannot be changed, and (b) in the sense that there is no time at which the state of affairs is non-existent. The assertion that a state of affairs that has obtained is a possible state of affairs is true only when 'necessity' is understood in the latter sense. In this latter sense, and as explained at the very beginning of this subsection, once a state

[10] Compare it with the following claim: 'A state of affairs that *has existed* is a possible state of affairs'. The grammar of the sentence (present perfect tense + a stative verb) leaves open the possibility that the state of affairs that the claim is about might be a necessary one.

of affairs obtains (notice, again, that 'obtain' is an action verb, which indicates that the state of affairs did not always exist), it is describable as a *possible* state of affairs.

Now, things get a bit complicated once we go beyond a merely conceptual reading of the assertion. One might wonder whether and in what sense the assertion *works* out there in the world. Does the assertion as it stands imply that, for instance, if a state of affairs has obtained, then it is possible that it will obtain again? Surely, one would be justified in reading it in this way. A moment's reflection reveals, however, that, at least for some states of affairs, this cannot be true. There are possible states of affairs that, (a), involve perishable particulars and, (b), involve events that cause the involved particulars to perish. Clearly, once such a possible state of affairs has obtained and the particulars involved have perished, it is impossible for that state of affairs to obtain again (that is, to obtain in future). Thus, there are states of affairs that have happened and yet are *in a sense* impossible. For reasons that will become obvious later, we need to be able to read the assertion as implying that if a state of affairs has obtained then it is possible it will obtain again. To allow that reading, the assertion needs to be qualified in something like the following way:

> AQ: If a state of affairs has obtained and if its relevant subject(s) has/have survived the obtaining, then it is possible that the state of affairs will happen again.

Not much turns on the notion of a relevant subject (of a state of affairs), and we do not intend to give a definition of it here. Its function in the qualified assertion is just to block the application of the assertion onto the states of affairs that happened but cannot possibly happen again because its relevant subject(s) has/have perished. Clearly, the death of Bertrand Russell implies that a state of affairs in which Bertrand Russell has died is, *conceptually* speaking, a possible state of affairs. It is not, however, a possible state of affairs in the sense of (AQ) as this state of affairs cannot happen again simply because the subject of this state of affairs does not exist any more.

We believe that our ordinary intuitions about possibility are governed by, among others, something like (AQ). Consider the case of climbing

Mount Everest without the use of supplemental oxygen. For a long time, it had been hotly disputed whether this was possible at all. Then, on 8 May 1978, Messner and Habeler reached the summit of Mount Everest without the aid of supplemental oxygen. This achievement has established that it is possible for man to climb Mount Everest without supplemental oxygen.[11] It has been established both *conceptually* and in the sense of (AQ). In the sense of (AQ), it has been established because it has happened and the subject of this kind of state of affairs—man—is still around to possibly repeat the performance.

3.3 Plausibly, if a state of affairs involving a member m of a kind k is a possible state of affairs then a relevantly similar state of affairs involving any other member of the kind k is possible too

We have assumed this in the last paragraph of the previous section where we treated a particular achievement of two Italian mountaineers as indicative of what is possible for man as a species. In a footnote related to that paragraph (footnote 11), we gave a brief consideration in support of this treatment.

Let us repeat and slightly expand the supporting consideration. There is a particular tree in California's White Mountains that has been named, quite tellingly, *Methuselah*. The tree belongs to the species of bristlecone pine trees and is believed to be almost 5,000 years old. This makes it the oldest non-clonal tree in the world. Now, as far as we know, no other bristlecone pine tree is as old as Methuselah, and the vast majority of the other bristlecone pine trees have lived nowhere close to 5,000 years. Yet, it is fairly common to generalize from what we know about Methuselah to what we take as possible about the species that Methuselah belongs to. When you start reading about these amazing trees, you will often come across

[11] Here we presume as unproblematic to generalize from a member of a species to species as a whole. It sounds very natural to say that a bristlecone pine tree can live (i.e. it is *possible* for it to live) for more than 5,000 years on the grounds that just one particular member of this species, Methuselah in California's White Mountains, has lived that long, even though vast majority of bristlecone pine trees have lived nowhere close that long. We will say more about this in the following section.

something like the following perfectly natural sounding statement: 'The bristlecone pine can live 5,000 years, making it the oldest individually growing organism on the planet, […] (Smith 2017). This statement is a good example of precisely the kind of a generalization from a single (past) achievement of a member of a species to what is possible for the species as a whole that we endorse here as plausible. The context of the article this statement is taken from makes it clear that 'the bristlecone pine' refers to a species, 'can' refers to possibility, and the figure of 5,000 years relates to Methuselah.

3.4. Some men did or have done things that (have) corrupted their moral standing

This is an uncontroversial empirical fact. The list of serious wrongdoings done by millions of men throughout the history is disturbingly long. In a suitable context, any of those serious wrongdoings would be deemed sufficient to critically undermine one's moral standing.

3.5. It is possible for any man to have their moral standing corrupted

This is step (5) of the argument, and it follows unproblematically from the previous three steps. Step (4), when slightly reformulated, says that some members of the species of Homo sapiens—that is, some men—have done things that have corrupted their moral standing. Step (2) says that a state of affairs that has obtained is a possible state of affairs. This then entails that a state of affairs in which a member of the species Homo sapiens does things that corrupt his/her moral standing is a possible state of affairs. Step (3) allows generalizing from a possibility about an individual member of a species to a possibility about the species as a whole. Hence, we can conclude that it is possible for any man to have their moral standing corrupted.

3.6. One's (present) moral standing gets corrupted (not only by the past and present wrongdoings but even) by one's future wrongdoings

On the face of it, this sounds rather unintuitive. How can one's present moral standing be corrupted by a future wrongdoing, i.e. by a wrongdoing

that has not happened yet? The unintuitiveness of the claim has to do with our common understanding of how causality and the arrow of time relate to each other. Normally, we will not take future events as being causally efficacious in the present because it seems a fundamental fact that the future is due to the present and not vice versa.[12] And it might seem like that's what we are being asked to do here: we are invited to consider and accept that one's *future* wrongdoing affects one's *present* moral standing. The appearance is misleading though. The relation between one's wrongdoing and one's moral standing is not a causal relation, or, at least, not a straightforwardly causal one. We do not wish to go into the metaphysics of causality here to illuminate the point. Instead, consider the following simple analogy: You have acquired insider information that in two months the government will introduce a drastic currency reform that will depreciate the value of the pound ten times. In response to this information you withdraw all the savings from your bank account and buy gold with it.

Now, there is a sense in which this little story could be described as a future event affecting your present state or actions: a currency reform happening in two months causes you to adopt (now) certain financial measures in response.[13] Thus, as we can see, there is a perfectly natural way of taking future events as affecting the present.

The analogy, however, will take us only so far. The target claim is that one's future wrongdoing corrupts one's present moral standing. In the case of the currency reform, its happening in the future—although it, in a sense, affected my *actions* in the present—has not done anything to the value of the *currency* in the present. Apart from a few insiders in the government

[12] This is not to say that this seemingly uncontroversial fact has not been challenged. There is some intriguing literature on the issue of backward causation that seriously discusses the possibility of cases where the effect temporally, but not causally, precedes its cause. See for instance (Dummett 1954) or (Faye 1997).

[13] This, of course, is neither the only nor the most natural reading of the little story. It could be insisted that it is rather a present *belief* about a future state/event than the future state/event itself that causally affects your actions in the story. And we would not want to object to this. Our point here is just that there is quite a natural way of taking future states/events as affecting the present, and that this little story is, by way of analogy, a first approximation towards understanding how one's *future* wrongdoing could be taken as corruptive of one's *present* moral standing.

and the Central Bank, no one knows about the planned reform, and, therefore, there are no bank runs that would depreciate the currency. Now, clearly, the currency and the moral standing are counterparts in the analogy. Thus, if we cannot conclude that the currency has devaluated then we cannot, by analogy, conclude that the moral standing has become corrupted. More needs to be done. Consider the following thought experiment: There is a time-machine device that makes it possible to find out what (if any) moral wrongdoing a person does in future. John's wife Mary has found out about John's extramarital affair. Mary reproaches John for cheating on her: 'You are a despicable person. How could you betray me like this?! You make me sick.' In response, John turns on the time-machine device. The device informs them that in 4 years from now Mary will also cheat on her husband. John breaks the silence: 'Who are you to call me a despicable person? You are no different!'.

Now, we believe it will be agreed that John's reply to Mary's reproach, and his questioning of her moral standing, is totally appropriate. It seems clear that Mary has it in her to cheat on her husband and that she deserves the resentment that John has expressed against her. And if that is so, then we can conclude that one's future wrongdoing, in some sense, corrupts one's (present) moral standing. At the same time, it should be noticed that the way one's future wrongdoing affects one's present status might not work for other future facts or actions. Consider, for instance, a non-swimmer about whom the time machine reports she will become, in the future, a competent swimmer. Surely, we would not want to claim that due to this *future* fact she is a competent swimmer *now*. Similarly, we all are going to die at some point in future, and yet we do not conclude that we are dead now. The two examples suggest there is something unique about (at least some) morally-laden facts/actions when compared to (at least some) other kinds of facts/actions. The little story of John and Mary seems to show that the former can affect the status of relevant agents in a way that the latter cannot[14].

[14] Notice also that a competent swimmer can, for whatever reason, become a non-swimmer. A murderer, on the other hand, will remain a murderer even if he never murders again. This again shows that morally laden facts/actions affect the status

3.7. (PP) can be applied to (5), that is: the status of one's moral standing belongs to the states of affairs that (PP) applies to

In section 3.2 we have identified an ambiguity in the notion of possibility. *Conceptually* speaking, a state of affairs that has obtained is a possible state of affairs. However, the same state of affairs will not be possible in the sense of it possibly obtaining again if it, (a), involves perishable particulars and, (b), the state of affairs is such that when it has obtained, the relevant particulars perished.

Now, in the following section, we will want to conclude that for all men there is a time at which their moral standing gets corrupted. We already know that it is possible for any man to have their moral standing corrupted. At the same time, many (perhaps most) men have not, yet, had their moral standing corrupted. This means that if we want to conclude that for all men there is a time at which their moral standing gets corrupted, then that time must be in the future. There is, however, something else waiting in the future for all men—their death. Any man is a perishable particular and it is certainly possible that they will perish before they manage to corrupt their moral standing. Surely, it would be extremely implausible to claim that none of the people presently alive will die before they manage to corrupt their moral standing. This possibility represents a serious challenge to our argument, because if it is possible for a man to perish before they corrupt their moral standing then we will not be able to conclude that for *all* men there is a time (in the future) at which their moral standing becomes corrupted.

One way to respond to this challenge is to appeal to certain implications that can be extracted from the following assertion:

L: *Luck* cannot make a difference in one's moral standing.

Nagel (1993, 58) refers to something like (L) when he says that it is 'intuitively plausible that people cannot be morally assessed for what is not their

---

of relevant agents across time spans in a way that other kinds of facts/actions cannot. It also shows that a moral wrongdoing has a permanent effect on one's status as it 'overrules' *the possibility* that one's moral status is, at the same time, uncorrupted.

fault, or for what is due to factors beyond their control'. Nagel claims this in the context of discussing what has been known as the Problem of Moral Luck (PML). (PML), roughly, is a problem constituted by an obvious tension between the intuitively compelling (L) and the fact that in our common practice of holding others morally responsible, luck does seem to make a difference (for instance, a drunk driver that ran over a pedestrian will be blamed more than a drunk driver who was lucky that there were no pedestrians around when he was driving home from the pub). We do not need to go into the intriguing details of (PML) here. For our purposes, it suffices to notice that (PML) is a problem taken seriously by contemporary moral theorists, which can be the case only if the moral intuition that co-constitutes it—that is, (L) - is taken as sufficiently plausible. We will follow suit. So how exactly does (L) help us to respond to the challenge mentioned above? There is an essential aspect of luck that could be described as a lack of control. This should be uncontroversial. Surely, an event that is under one's control cannot be described as a *lucky* event. If that is so, then (L) could be reformulated in the following way:

> L*: An event that one has no control over cannot make a difference in one's moral standing.

We take it that (L*) is no less plausible than (L).[15] Now, one's mortality is clearly beyond one's control and as such it belongs to the kind of events that cannot make a difference in one's moral standing. In other words, death is not an *excuse* for wrongdoings that one would do if one did not die. Consider this:

A terrorist plants a bomb in a theatre full of people. The bomb is controlled remotely. The terrorist contacts the authorities informing them that a bomb has been planted in an unknown public place. He presents a list of

[15] However, we do not wish to imply that *luck* and *lack of control* are synonymous concepts. They are not. There is going to be full moon in several days; an event that is totally beyond my control. To describe this as me being lucky (whenever there is full moon) would be rather weird. There is remarkably little discussion on the nature of the concept of luck among moral theorists. An intriguing exception is (Rescher 1995).

demands. After the negotiations with the authorities fail, the terrorist proceeds to detonate the bomb. He is about to push the button on the remote control when he suffers a sudden heart attack. He drops the remote control before he manages to press the button; he passes out and a few minutes later he is dead.

It will be agreed, we believe, that the terrorist in our little story is an extremely wicked person. He will be seen as such despite the fact that he has caused no harm to anyone. He would have done it, had he not been prevented by his sudden death, and that is enough for us to judge him as morally corrupted.[16]

We can conclude that in the contexts of evaluating moral standings, one's perishability is irrelevant. That is, in these contexts, a human agent must be seen as if she was imperishable. Above (section 3.2.), it has been shown that (PP) will not work for perishables. Thus, if man can be treated as imperishable in the contexts of evaluating their moral standing, then (PP) applies in those contexts.

3.8. For all men there is a time at which they do things that corrupt their moral standing

This follows from steps (1), (5) and (7) in the following straightforward way. (5) tells us that it is possible for any man to do things that corrupt their moral standing. (1) tells us that for some states of affairs, if they are possible then there is a time at which they obtain. And (7) tells us that doing things that corrupt one's moral standing belongs to the state of affairs that (1) applies to. Step (7) is crucial here, as it allows us to treat man as immortal. Without this step it could be objected that many men will simply die *before* they manage to do something that corrupts their moral stranding.

[16] It could be pointed out that what makes us judge the terrorist as morally corrupted is his *intention* to do harm. This could then be taken as showing that it is not only future (or past or present) actions that corrupt one's moral standing but the *intentions* to do them as well. This can be conceded without any harm to the logic of the argument. The reader is invited to understand an action as morally corrupting only if it is intentional.

3.9. The moral standing of man is essentially corrupted

Above, we have concluded that for all men there is a time at which they do things that corrupt their moral standing. It seems obvious that a wrongdoing done in the past or in the present corrupts one's *present* moral standing. It is much less obvious, however, that one's *present* moral standing gets corrupted by a *future* wrongdoing. Step (6) explains how that is the case nonetheless.

At this point we can draw something like the following picture of man's moral standing. It is possible for any man to do a wrongdoing. There is a time at which all men will do a wrongdoing. Regardless of whether the time of one's wrongdoing is in the past, in the present or in the future, the wrongdoing corrupts one's (present) moral standing. Therefore, at any present moment, all men's moral standing is corrupted. Another way of putting the last point is that: the moral standing of man is essentially corrupted.

Let us briefly return to an objection that we have touched upon above (section 3.6.). The objection: 'There is a time at which all men will die' is (in the context of our discussion here) analogous to 'There is a time at which all men will do wrongdoing'. From 'There is a time at which all men will do wrongdoing' we have concluded that 'Therefore, at any present moment, all men's moral standing is corrupted'. The logic behind this step seems to commit us to endorsing an analogous step from 'There is a time at which all men will die' to 'Therefore, at any present time, all men are dead'. This, of course, is a false conclusion, which casts serious doubts on the validity of our central argument. At the end of section 3.6., we gesture towards a reply to this objection. Let us slightly expand on the reply. The little story of John and Mary shows that it feels quite natural to *now* see a person as morally tainted on the grounds of the person's *future* wrongdoing. It will be noticed that the little story (once some of its inconsequential details has been changed) will not elicit the same intuition if instead of being about one's moral wrongdoing it is about one's death. Consider the following conversation:

John's wife Mary has found out that John killed their neighbour, Paul. 'You murderer!', Mary screams at John. John calmly replies, 'Inevitably, Paul was going to die at some point in future. This implies that Paul was

dead before what you seem to see as his 'murder'. Surely, you cannot murder a dead man'.

Paul's reply, unlike the one he gives in the original story, has no force. There is, let's assume, incontrovertible evidence that Paul was alive before John killed him, and the fact that Paul was certain to die at some point in future does not intuitively feel to have any bearing whatsoever on Paul's well-being prior to John's horrendous crime. We could speculate here about what exactly it is about moral wrongdoings that they can (unlike one's death and most other actions/events) affect, in some relevant way, ones's present status from the future. We will not do it here though as we can generate enough support for the claim (that one's *future* wrongdoing corrupts one's *present* moral standing) from the intuitive force of the little story about Mary's future wrongdoing.

4. Conclusion

Our moral intuition dictates that a morally compromised person should not hold others morally accountable. Above we have argued that, in fact, we all are morally compromised. This implies that our common practice of holding others morally accountable is always inappropriate.

The argument relies heavily on the statistical interpretation of modality and a thought experiment designed to show that one's future wrongdoing can affect one's present moral status. Even though the statistical interpretation of modality is not popular among contemporary philosophers and logicians, it rests on quite an intuitive conception of possibility and necessity. On the face of it, the statistical approach to modality seems to commit its advocates to some rather unpalatable implications. We have addressed the implications above showing, as we believe, that the statistical approach to modality is a defensible position.

The claim that we all are morally compromised and as such should refrain from holding others morally accountable could feel rather extreme. The following two brief comments might make it more plausible. First, the claim that we all are morally compromised is not a new one. It has a venerable precedent in the theological notion of original sin that has shaped

Christian ethics. Second, we do not, in fact, encourage abolishing the practice of holding others morally accountable. We believe there are reasons—psychological and utilitarian—for preserving the practice. We only wish the claim to be understood as a humbling reminder that the practice is not morally innocent.

References

Barnes, Jonathan. 1977. Review of "Time and Necessity: Studies in Aristotle's Theory of Modality," by Jaakko Hintikka. *The Journal of Hellenic Studies* 97: 183–6. https://doi.org/10.2307/631049

Becker, Oscar. 1952. *Untersuchungen über den Modalkalkül*. Meisenheim am Glan: Anton Hain.

Beebee, Helen. 2006. "Does Anything Hold the Universe Together?" *Synthese* 149 (3): 509–33. https://doi.org/10.1007/s11229-005-0576-2

Coates, Justin D., and Neal A. Tognazzini, eds. 2013. *Blame: Its Nature and Norms*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199860821.001.0001

Dummett, Michael. 1954. "Can an Effect Precede its Cause?" *Proceedings of the Aristotelian Society* 28 (Supplement): 27–44. https://doi.org/10.1093/aristoteliansupp/28.1.27

Faye, Jan. 1997. "Causation, Reversibility, and the Direction of Time", In *Perspectives on Time*, edited by Jan Faye, Uwe Scheffler and Max Urchs, 237-266. Dordrecht: Kluwer Academic Publisher. https://doi.org/10.1007/978-94-015-8875-1_11

Friedman, Marylin. 2013. "How to Blame People Responsibly." *The Journal of Value Inquiry* 47 (3): 271–84. https://doi.org/10.1007/s10790-013-9377-x

Hintikka, Jaakko. 1976. "Kant on 'The Great Chain of Being' or the Eventual Realization of All Possibilities: A Comparative Study.' *Philosophic Exchange* 7 (1): 69–89.

Hintikka, Jaakko. 1980. "Aristotle on the Realization of Possibilities in Time." In *Reforging The Great Chain of Being*, edited by Simo Knuuttila, 57–72. Dordrecht: Springer-Science+Business Media,B.V. https://doi.org/10.1007/978-94-015-7662-8_3

Kant, Immanuel. 1966. *Critigue of Pure Reason*. Translated by F. Max Müller. Garden City, N.Y.: Doubleday.

Lovejoy, Arthur. 1936. *The Great Chain of Being*. Cambridge, Massachusetts: Harvard University Press.

Nagel, Thomas. 1993. "Moral Luck." In *Moral Luck*, edited by Daniel Statman. 57–71. New York: State University of New York Press.

Rescher, Nicholas. 1995. *Luck: The Brilliant Randomness of Everyday Life*. New York: Farrar, Straus and Girou).

Russell, Bertrand. 1956. "The Philosophy of Logical Atomism." In *Logic and Knowledge*. Essays 1901–1950, edited by R.C.Marsh. London: Allen & Unwin.

Smith, Scott. 2017. "Scientists: The Future of Oldest Tree Species on Earth in Peril." AP News. Accessed June 26, 2020. <https://ap-news.com/776e453d15674f1e9eb20af289d6e46e/Scientists:-Future-of-oldest-tree-species-on-Earth-in-peril>

Strawson, Galen. 1989. *The Secret Connexion: Causation, Realism, and David Hume*. Oxford: Clarendon Press.

Organon F 29 (3) 2022: 376–388
https://doi.org/10.31577/orgf.2022.29304

ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

Concepts May Still Be Objects

Harold Noonan*

Received: 11 May 2022 / Revised: 20 August 2022 / Accepted: 28 August 2022

Abstract: In his book (2021) Trueman attempts to provide a solution to the problem of the concept horse, which according to Frege's published writings is an object, not a concept. In the course of doing so Trueman rejects Wright's response (1998) according to which some objects are also concepts, for example, the concept horse, so the categories are not exclusive. Trueman's argument for exclusivity (Chapter 4) is the heart of the book, and as he says , it is his response to holders of differing views, like Wright. I think that there is a gap in Trueman's argument which needs to be filled if Wright is to be considered refuted.

Keywords: Concept; objet; Wright; singular term; predicate; Frege.

Concepts may still be objects

In his book (2021) Trueman attempts to provide a solution to the paradox of the concept *horse*, which according to Frege's published writings is an object, not a concept. In the course of doing so Trueman rejects Wright's response (1998, 2001) according to which some objects are also concepts,

* University of Nottingham
https://orcid.org/0000-0001-5538-5444
University of Nottingham, Department of Philosophy, Room C41 Humanities, University Park, Nottingham, NG7 2RD, United Kingdom.
Harold.Noonan@nottingham.ac.uk

for example, the concept *horse*, so the categories are not exclusive. According to Wright this is so because the concept *horse* is an object in virtue of being the *reference* of the singular term 'the concept *horse*' and a concept in virtue of being the *ascriptum* of, what is *ascribed* by, the predicate 'is a horse'. Reference and ascription are two different relations, but some things, concepts, are both referents and ascripta. Wright's aim is to offer a resolution of the paradox of the concept horse whilst retaining the intuitive sounding reference principle, that sameness of reference ensures sameness of semantic role, which making the distinction between reference and ascription allows him to do.[1] His motivation is to allow that a distinguished category of entity can be associated with predication, as objects (or particulars) are associated with the use of singular terms, but to avoid paradox by distinguishing the relation between predicate and its associated entity as a sui generis one.

Trueman's argument for exclusivity (Chapter 4) is the heart of the book, and as he says (2021:98), it is his response to holders of differing views, like Wright. I think that there is a gap in Trueman's argument which needs to be filled if Wright is to be considered refuted.

Trueman argues that we have to recognize two notions of reference, term-reference and predicate-reference. Singular terms term-refer, predicates predicate-refer. Term-referents are objects, predicate-referents are concepts, though as Trueman explains, this is loose talk (see e.g., 2021, 55, fn. 13). Predicate-reference is the relation between a predicate and a concept which is analogous semantically to the first-level relation of term-reference in which singular terms stand to objects.

A statement of term-reference takes the form of, for example:

(1) 'Socrates' refers to Socrates.

[1] Wright (2001, 72) notes at the beginning of the paper that his solution may however be subject to other well-known problems such as the intensional version of Russell's paradox. But at the end of the paper, noting this again, he writes (2001, 90): 'this, like the recent resurgence of tuberculosis in the western world, is a disappointment. But I do not think it is really an objection — too many of the family of paradoxes that exercised Russell survive the imposition of Frege's hierarchy to allow us to think that it gets to the root of that particular one.'

In this 'Socrates' occurs twice, once mentioned, once used—the statement disquotes. And it is reasonable to think that this is what someone knows if he understands the meaning of 'Socrates' (which he may do, of course, without understanding the use of quotation marks, or the meaning of 'refers' in English, or use any word at all which means 'refers').

In parallel, a statement of predicate-reference must disquote. It must take the form of, for example:

(2) 'is wise' predicate-refers to is wise.

Of course, this looks ungrammatical. But it is not, because 'predicate-refers to' is a novel expression we can use how we like. So (2) is grammatical by stipulation. But what does it mean? Trueman (2021, 71) says that for his purposes we can take it to mean: 'is wise' is true of z iff z is wise (or: z satisfies 'is wise' iff z is wise). It is reasonable to think that this is what someone knows if he understands the meaning of 'is wise' (which he may do, of course, without understanding the use of quotation marks, or the meaning of 'is true of' in English or use any expression at all which means 'is true of').

Now let us shift to another example of term-reference:

(3) 'The property (concept) of being wise' term-refers to the property of being wise.

Now you might think that conjoining (3) and (2):

(3)+(2) 'the property of being wise' term-refers to the property of being wise and 'is wise' predicate-refers to is wise

– you get a *witness* to the existential generalization:

(EG1) for some x, x is an object and x is a concept

i.e.,

(EG1*) for some x, there is a term which term-refers to x and a predicate which predicate-refers to x.

But you don't, Trueman (2021, 51-2 and 60) points out, since what follows 'term-refers to' in (3)+(2) is different from what follows 'predicate-refers to'; one is a term and one is a predicate.

Similarly, (3)+(2) is not a witness to the second-order existential generalization:

(EG2) for some Y, Y is an object and Y is a concept.

So Trueman concludes that the statement that there is something which is both an object and a concept is nonsensical since it has not only no true instance, but no possible instance since terms are not predicates. (He explains (2021, 117) what he means by saying that this is nonsense: The meanings of the words in the sentence 'there is something which is both an object and a concept', as he understands them, are such that there is no meaningful sentence composed of words with those meanings arranged in that way.)

Trueman's thought that there is no single relation of reference in which both terms and predicates can be regarded as standing may be Frege's. It is certainly entirely in accord with Frege's thought and emphasized by Dummett (1973, 182–83, 171, 411), as Oliver (2005) points out. But it is not disputed by Wright. In fact, as noted, it is a key point in his article. However, Wright contends that some objects are concepts (e.g., the property (concept) of being wise). So how can Wright reply? How can he produce a witness to (EG1)?

I think that there is only one way, which appeals to the Fregean thought that the same content can be carved up in more than one way.[2]

First, a simple example of this. Consider 'John hates Mary's father'. This asserts a relation between John and Mary—he hates her father. It also asserts a relation between John and Mary's father (who may be Peter, or John himself if John hates himself)—he hates him. We can carve up the content in (at least) two ways. Frege's example is 'the direction of line a is identical with the direction of line b', as explained by the equivalence with 'a is parallel to b' (the Fregean 'direction principle'). We can regard this

[2] For a statement by Wright of this Fregean doctrine see Hale and Wright (2012: 120-1): 'We owe to Frege the insight that one and the same thought may allow of decomposition into distinct logical forms: the thought that Socrates is wise, for instance, may be regarded both as a first-level predication of wisdom of Socrates, and as a second-level predication, of "Socratising", or applying to Socrates, of wisdom.'

as asserting identity between the direction of line a and the direction of line b or as parallelism between a and b.

Now let us consider the thought that (2), 'is wise' predicate-refers to is wise, i.e.,

'is wise' is true of something iff it is wise.

And let us express this, using Wright's terminology, as:

(WL) 'is wise' ascribes the property which something possesses/the concept something falls under iff it is wise.

No one can prevent us doing this (just as no one could prevent Frege stipulating that 'the direction of line a is the direction of line b' is to be understood as expressing the thought that a is parallel to b). But this is a bit long, so we can abbreviate harmlessly to reach the shorter:

(WS) 'is wise' ascribes the property of being wise.

Now (WL) is true iff 'is wise' stands in a relation—predicate-reference—to what 'is wise' predicate-refers to. But it also, if we carve the content differently, is true iff 'is wise' stands in a different relation—ascription—to the term-referent of 'the property which something possesses iff it is wise', i.e., the term-referent of 'the property of being wise'. But can we carve the content differently? Well, one of Wright's long-held beliefs, also frequently argued for by Bob Hale (see e.g., Hale 2013, also Hale and Wright 2012, 113–14) is the syntactic priority thesis, the thesis that Fregean objects are to be understood as the (term)-referents of singular terms, and singular terms are to be understood just as expressions which behave as singular terms. But we can recognise 'the property which something possesses if and only if it is wise' in (WL) as a singular term, and so as standing for an object if anything. And if we do, we must recognize (WL) as asserting, as well as the relation of predicate-reference, the relation of ascription between the predicate 'is wise' and that object. So we must recognise its content as capable of being carved up in different ways. And so recognising it, it is to be noted, we do not have to deny the essential role of knowledge of the truth of a thought expressed by a statement in which the predicate is disquoted to our understanding of the predicate 'is wise' since in (WL) 'is wise' occurs both quoted and not quoted (and (WS) is only intelligible as an

abbreviation). Hence Trueman's requirement that the statement that relates to the predicate 'is wise' as (1) relates to the name 'Socrates' must disquote, as (1) does, is satisfied, on Wright's account.

But then (3) in conjunction with (WS) *is* an instance of, a witness to the truth of:

(EG1) for some x, x is an object and x is a concept

– if we understand 'x is an object' to mean 'there is a term which term-refers to x' and 'x is a concept' to mean 'there is a predicate which *ascribes* x', since 'the property of being wise' does term-refer to the property of being wise, and the predicate 'is wise' does ascribe the property of being wise. (Note that on this explanation 'concept' is a genuine count-noun applicable to objects, so its meaning so understood is different from the meaning it has on Trueman's account;[3] when Wright says 'concepts are objects' he is not

[3] Trueman does not have to give 'concept' any meaning at all, but he notes that it is useful to use sentences in which it occurs to make certain claims easier to pronounce, e.g., to use 'the concept wise is a concept' to mean 'for some F, for every x, Fx iff x is wise' (2021, 112), or to use 'it is nonsense to say that a concept is an object' to express the view ascribed to him above about the *sentence* 'there is something which is both an object and a concept'. (Note that we can also use the apparently object-language sentence 'nothing is both an object and a concept' to express this view (2021, 215). It remains that if this sentence is *not* understood in this way it must either say something irrelevant, whether true or false, to Frege's concerns in 'On Concept and Object'—e.g., that nothing is both an object and an extension—or be meaningless. There is no truth here that cannot be said, but only shown.) Trueman could also use the word 'concept' as part of a second-level predicate, 'Concept x, …x…', so defined that 'Concept x, Fx' means, say, 'for all x, Fx iff Fx'. He could do the same with other words denoting properties of concepts, or more generally of functions, including 'function' (specifically 'first-level function of one argument', and so on) itself (cf. Frege 1980, 141). So he could use 'Function x, …x…' so defined that 'Function x, Fx' means 'for all x, for some y, Fx=y', e.g., 'Function x, sin(x)' means 'for all x, for some y, sin(x)=y'. He could then use 'the sine function is a function' to assert this in a more easily pronounceable way. Frege himself renders 'The real function Φ(x) of a real variable x is continuous throughout the interval from A to B' (note the definite article 'The real …') in his logical notation (Frege 1979: 24). He adds: 'If in this case the formula seems longwinded by comparison with

asserting what Trueman denies when he says (speaking loosely) 'concepts are not objects'.)

So Wright is not refuted unless it can be shown that the thought that:

(WL) 'is wise' ascribes the property which something possesses iff it is wise,

the thought that 'is wise' is true of something iff it is wise, cannot be understood as asserting two different relations to obtain, one of predicate-reference, one of ascription, one of unequal level, R(x,Y), one of first-level, R(x,y), as according to Frege, 'the direction of a is parallel to the direction of b' can be understood as asserting both identity and parallelism (and as, evidently, 'John hates Mary's father' may be understood as asserting two different relations to obtain). But why not? Or is Frege wrong too?

In fact, it may be that Wright, writing with Hale (2012, 117–19), has anticipated and answered Trueman's argument in a response to Noonan (2006).

Wright and Hale write:

> Noonan claims … a fatal incoherence in Wright's claim that a concept can be both *ascriptum* of a predicate and reference of a singular term:
>
> > ...the expression 'ascription' was coined [by Wright] to indicate just that unequal-level relation between a predicate and a concept which is analogous semantically to the first-level relation between singular term and referent. It is, in fact, that relation expressed by: … applies to something if and only if it … (e.g. 'is a horse' applies to something iff it is a horse)
>
> But if [so], Noonan argues, Wright's position is incoherent. For he must hold that … there is some X such that 'is a horse' ascribes X and 'the concept *horse*' refers to X. But '"is a horse" ascribes X' expands as '"is a horse" applies to something iff it X', so Wright needs to hold that some instance of the schema:

the verbal expression, you must always bear in mind that it gives the definition of a concept which the latter only names.'

> > ‘is a horse’ applies to something iff it X and ‘the concept horse’ refers to X
> >
> > is true. But there can be no true instance, … since what must replace the first occurrence of ‘X’ is a predicate, … the second a singular term.

Hale and Wright go on:

> > But [Noonan’s] objection is mistaken. Wright’s explanation [of ascription] runs as follows:
> >
> > > For a predicate to stand in the relation of ascription to a property or concept is just this: for its sense so to relate it to that property/concept that it may be used in concatenation with an appropriate singular term to say of the bearer of that term that it has the property, or falls under the concept in question
> >
> > Wright does not explain ‘ascribes’ as an *unequal-level* relational expression.... Noonan is just not taking seriously Wright’s view that concepts/properties are a kind of object. Given that they are a kind of object, there is no reason why the relation of ascription should not be a relation expressed by a first-level predicate…. Thus Wright *can* say that:
> >
> > ∃x(‘is a horse’ ascribes x and ‘the concept horse’ stands for x)
> >
> > This has a true substitution instance:
> >
> > > ‘is a horse’ ascribes the property of being a horse and ‘the concept horse’ stands for the property of being a horse. (Hale and Wright 2012, 117–19)

Thus, Hale and Wright argue, Wright *can* give an instance of the claim that something is both a concept and an object (as Wright understands these terms, i.e., as the ascriptum of a predicate and the referent of singular term).

If Fregean recarving of content is permissible in this case, I think that Noonan, and along with him Trueman, is refuted. That is, it has *not* been shown by either that there is no equal-level relation expressible by a first-level predicate satisfying the description by which Wright explains ‘ascription’.

I elaborate. Hale and Wright's key objection to Noonan's objection to Wright's original paper is that he simply ignores Wright's explanation of ascription. Wright's view, as Noonan takes for granted, is that knowledge of the concept a predicate ascribes is constitutive of understanding—in the Fregean terms he uses, grasping the sense of—the predicate. But knowing what the predicate is true of (applies to)[4] is what is constitutive of this understanding. So knowledge of the concept ascribed *is* knowledge of what the predicate is true of.[5] (If it were not, one could not inform someone, in German say, that 'is wise' ascribes the concept wise and thereby provide him with the knowledge needed to use the predicate comprehendingly.) Noonan takes it to follow that Wright's ascription must be the unequal-level relation Trueman calls 'predicate-reference'.

Hale and Wright's point is that this does not follow from Wright's view that knowledge of the concept a predicate ascribes is knowledge of what it is true of. Wright's explanation of ascription is in accordance with this view, but it identifies ascription as a relation expressible by a first-level predicate. According to it, 'is wise' ascribes the concept wise/the property of being wise just in case its sense so relates it to that concept that it may be used in concatenation with an appropriate singular term to say that the bearer of the term falls under the concept, so that grasp of the sense of the predicate (understanding it) requires knowledge that concatenation of 'is wise' with an appropriate singular term—'N is wise'—says that the bearer of 'N' falls under the concept wise. But given that objects are just the bearers of singular terms, that to which they refer, this is merely to say that 'is wise' ascribes the concept wise just in case its sense so relates it to that concept that understanding it requires knowing that 'is wise' is true of an object if and only if it falls under the concept wise. That is to say, so relates it to that concept that understanding it requires knowledge that 'is wise' is true of an object if and only if it falls under the concept something falls under just in case it is wise ('the concept wise' is merely an abbreviation). But

[4] I.e., knowing the thought, in the case of the predicate 'is wise', expressed by a sentence of the form: 'is wise' is true of something iff it is wise.

[5] Cf. Hale (2013, 35–6): 'a predicate stands for a certain property … knowing what property the predicate stands for will just consist in understanding the predicate, i.e., knowing what something must be like, if the predicate is to be true of it'.

something falls under the concept something falls under if and only if it is wise if and only if it is wise (or: possesses the property something possesses if and only if it is wise if and only if it is wise). So 'is wise' ascribes the concept wise iff its sense so relates it to that concept that understanding it requires knowing that 'is wise' is true of an object iff it is wise. So (WL) (= (2), given Trueman's definition of 'predicate-refers') is implicit in Wright's explanation of ascription and Hale and Wright's point is that the content of this can be carved, not only in a way that reveals it as asserting an unequal-level relation (predicate-reference), which Wright of course accepts,[6] but also in a way that reveals it as asserting a relation expressed by a first-level predicate (ascription).[7]

This is not to say that Hale and Wright's defence of concepts as objects is unassailable. It is essential to that position that the content of 'x falls under the concept something falls under if and only if it is wise' can be equated with that of 'x is wise'. But their position also requires that the former contains, as it were, two unknowns, 'concept' and 'falls under'; it is not merely a stylistic variant. But a single equation with two unknowns cannot be solved (what are the values of x and y if 2x+y=7?). So, by what route can someone innocent of the meanings of these *two* expressions in Hale and Wright's writings come to an understanding of them? Wright's explanation of ascription employs both. (In Frege's direction equivalence 'the direction of line a is identical with the direction of line b iff a is parallel to b' it is assumed that the meaning of 'is identical with' on the LHS is the

[6] Wright himself draws attention to the possibility of using 'refers' as Trueman uses 'predicate-refers': 'that is not to say that we could not stipulate that "refers to" should have a use linking the name of a predicate to an expression—par excellence, the predicate itself—for its semantic value. In that case, "'is a horse' refers to is a horse"—proposal (i) above—would be well-formed, but—just for that reason—"refers to", so used, would not speak of the relation that holds between a singular term and the object for which it stands' (2001, 85).

[7] This fits in with Wright's comment: 'what fits an object for ascription—what makes it a concept/property—will be that it is the referent of an expression formed by a certain kind of abstraction on a corresponding predicate. That will not be the situation of the ordinary run of singular terms, though of course the details remain to be worked through' (2001, 90, fn. 18). Tthe Direction Principle is, of course, the paradigm abstraction principle.

relation of numerical identity, so there is only one 'unknown', 'the direction of'.)[8] There is also the worry that Wright's explanation of 'ascription' is inadequate to the generality of the problem (Hale 2013, 30). Concepts are a special case of functions. The problem of reference to concepts is a special case of the problem of reference to functions. Ascription is just a special case of the relation—which is not, of course, reference—between a functor and the associated function (call it 'specification'). Now how are we to fill in the gaps in: 'For a functor to stand in the relation of specification to a function is just this: for its sense so to relate it to that function that it may be used in concatenation with an appropriate singular term to …. the bearer of that term that it … the function in question'?

It should be noted that Noonan also articulates a separate worry about Wright's position (Noonan 2006, 169-70), perhaps related to the first in the last paragraph. There is according to Frege (before Russell's Paradox), an object which is the term-referent of 'the extension of the concept horse', namely, the extension of the concept horse, and this, Frege suggests, may permissibly[9] be taken as the term-referent of 'the concept horse'. And, of course, there is a first-level relation between the predicate (which is an object) 'is a horse' and this object. It is the relation which is the predicate-reference of 'is true of something iff it is a member of' (taking extensions to be sets). Noonan uses 'ascribes$_E$' as an abbreviation for this. Now Noonan's worry is that nothing in Wright's explanation of ascription distinguishes it

[8] Perhaps the explanation should be understood as two-step. First, we have (a paradox-proof variant of) the abstraction principle: 'The concept which something falls under iff it is F = the concept something falls under iff it is G just in case all and only all Fs are Gs' and then a principle introducing ascription: 'is an F' ascribes the concept something falls under iff it is H just in case 'is an F' is true of something iff it is H'. An instance of the first fixes the reference of 'the concept something falls under iff is it a horse' ('the concept horse') and the second fixes the ascriptum (?) of the relational predicate 'ascribes'.

[9] Noting the presence of the definite article. 'If [Kerry] thinks … that I have identified concept and extension of concept, he is mistaken. I merely expressed my view that in the expression 'the number that applies to the concept F is the extension of the concept *like numbered to the concept F*' the words 'extension of the concept' could be replaced by 'concept'. Notice carefully that here the word 'concept' is combined with the definite article' (1969:48).

from ascription$_E$. If so, he claims, Wright has not succeeded in setting out a position opposed to Frege's, as he intends. It may be that this points to a genuine problem with Wright's proposal.[10]

Setting this separate worry aside, however, the main point of this note has been to suggest that unless content carving is either in general illegitimate, or illegitimate in this particular case, there is a gap in Trueman's argument, as there is in Noonan's, understood as an attempted refutation of Wright.

References

Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 2020. *Analytic Narratives*. Princeton: Princeton University Press.

Boncompagni, Anna. 2016. *Wittgenstein and Pragmatism: On Certainty in the Light of Peirce and James*. London: Palgrave Macmillan.

Bruner, Jerome. 2004. "Life as Narrative." *Social Research* 71 (3): 691–710.

Burbules, Nicholas C. 2017. "Wittgenstein's metaphors and his pedagogical philosophy." In *A Companion to Wittgenstein on Education: Pedagogical Investigations*, edited by Michael A. Peters, and Jeff Stickney, 123–33. Singapore: Springer. https://doi.org/10.1007/978-981-10-3136-6_8

Danto, Arthur. 1965. *Analytic Philosophy of History*. New York: Cambridge University Press.

Eisele, Thomas D. 1990. Wittgenstein's Instructive Narratives: Leaving the Lessons Latent. *Journal of Legal Education* 40 (1/2): 77–95.

Garrison, Jim. 2017. "Wittgenstein and Classical Pragmatism." In *A Companion to Wittgenstein on Education: Pedagogical Investigations*, edited by Michael A. Peters, and Jeff Stickney, 321–38. Singapore: Springer. https://doi.org/10.1007/978-981-10-3136-6_21

[10] The question is 'are Wrightian concepts Fregean extensions?' if the answer is 'No' what distinguishes these objects? Not their principles of individuation since concepts for Frege, and hence for Wright, are individuated extensionally (unlike properties on some current uses of the term, on which, e.g., the property of being a cordate is not the property of being a renate). So what else? But if Wrightian concepts are Fregean extensions what is the disagreement between him and Frege, that is, what is *the thought* to which he ascribes the truth-value True and Frege ascribes the truth-value False?

Gasking, Douglas Aidan Trist, and Alan Cameron Jackson. 1967. "Wittgenstein as a Teacher". In *Ludwig Wittgenstein: The Man and His Philosophy*, edited by K. T. Fann, 49–55. New York: Dell Publishing Co.

Gebauer, Gunter. 2017. The Turn to Anthropology. In *Wittgenstein's Anthropological Philosophy. History of Analytic Philosophy*, 75–107. Cham: Palgrave Macmillan https://doi.org/10.1007/978-3-319-56151-6_4

Gibson, Arthur, and Niamh Ann O'Mahony, eds. 2020. *Ludwig Wittgenstein: Dictating Philosophy: To Francis Skinner—The Wittgenstein-Skinner Manuscripts*. Cham: Springer Nature.

Kačerauskas, Tomas. 1999. "Is L. Wittgenstein Analytic or Hermeneutic Philosopher?" *Problemos*, 560: 71–84. https://doi.org/10.15388/Problemos.1999.56.6863

Kuusela, Oskari, Mihai Ometiță, and Timur Uc¸an. 2020. *Wittgenstein and phenomenology*. Abingdon, Oxon; New York, NY: Routledge.

McDonald, Henry. 2001. Wittgenstein, Narrative Theory and Cultural Studies. *Telos* 121: 11–53.

Rorty, Richard. 1989. *Contingency, Irony, and Solidarity*. Cambridge: Cambridge University Press.

Synytsia, Andrii. 2020. "Pragmatic Aspect of Wittgenstein's Philosophy of Education: A Modern Interpretation." *Journal of Education Culture and Society* 11 (2): 39–51. https://doi.org/10.15503/jecs2020.2.39.51

Unger, Peter. 2014. *Empty Ideas: A Critique of Analytic Philosophy*. New York: Oxford University Press.

Wittgenstein, Ludwig. 1922. *Tractatus Logico-Philosophicus*. Translated by Charles Kay Ogden. London: Kegan Paul, Trench, Trubner & Co.

Wittgenstein, Ludwig. 1968. *Philosophical Investigations*. Translated by Gertrude Elizabeth Margaret Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig. 1969. *On Certainty*. Translated by Denis Paul, and Gertrude Elizabeth Margaret Anscombe. Oxford: Blackwell.

Wittgenstein, Ludwig. 1993. Remarks on Frazer's Golden Bough. In *Ludwig Wittgenstein. Philosophical Occasions, 1912–1951*, edited by James Klagge, and Alfred Nordmann, 118–155. Indianapolis: Hackett Pub Co Inc. (in German and English).

Wittgenstein, Ludwig. 2006. *Culture and Value*. Translated by Peter Winch. Chicago: Chicago University Press.

Organon F 29 (3) 2022: 389–404
https://doi.org/10.31577/orgf.2022.29305
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

DISCUSSION NOTE

Value and Freedom

A review of P.M.S. Hacker's *The Moral Powers: A Study of Human Nature* (volume 4). John Wiley & Sons, 2021

Joanna Moncrieff*

1. What is morality

Like the other volumes in Peter Hacker's impressive series on human nature, volume four offers his wisdom on many subjects and is full of insights and penetrating clarity. Along with an analysis of the nature of morality, he covers the subject of evil, freedom and determinism, the science of happiness and, not to be forgotten, the meaning of life! This review will focus on the most fundamental points: that value is inherent in the nature of life, and that morality is a corollary of human nature and the world we live in that presupposes freedom of action. There is much more that is interesting and enlightening in this volume, however, and I encourage readers to read it and see for themselves.

The mainstream view, according to Hacker, regards morality as a realm that is independent of the material world of facts, including empirical facts about human nature. For much of human history morality was believed to be derived from God, or the Gods, and more recently (and occasionally in the past) from abstract, universal moral laws or principles. As a consequence, many philosophers have been puzzled about the existence of 'value

* University College London
https://orcid.org/0000-0003-1214-6974
Division of Psychiatry, University College London, London.
j.moncrieff@ucl.ac.uk

in a world of fact' (p. 8), and have treated morality as something both detached and mysterious.

Hacker takes a different view. As he highlighted in the first volume of Human Nature, ideas of value arise in the context of the living world; the 'notion of the good of a being is biologically rooted' (p. 175) (Hacker 2010). Living creatures are different from inanimate objects by virtue of having a life cycle during which they grow and develop, reproduce and eventually decay and die. Our understanding of life is framed by the normal life cycle of the organism and the activities that are typical of its species at various points of this cycle. Living beings can flourish and prosper or decline and suffer illness, loss of powers and death. We judge whether situations are good or bad for an organism in terms of whether they enable it to survive, mature, reproduce and function in an optimal way. 'All values arise from life' as Hacker puts it (p. 7).

It is against the background of facts about the nature of human beings and of the environment we live in that we need to understand morality. Hacker points out how we have innate tendencies both to competitiveness and aggression, and to sympathy and cooperation. We experience sexual attraction and desire. We are born immature and need nurturing, and we are susceptible to illness, injury and death. We have the ability to emulate and to learn. We have the capacity for language and the ability to reason. We can follow rules and pursue goals. We have a range of emotions, both animal emotions (e.g. fear, affection and curiosity) and specifically human ones that depend on our mastery of language (e.g. pride and shame), which reflect that we care about things.

Human beings are also social creatures. We are dependent on each other for survival and for the realisation of many of our capacities, including our unique capacities for language and reasoning. Most humans also enjoy and seek relationships with others, as lovers or friends. Human morality is explicable with reference to these 'powers and propensities' that characterise human beings, including basic biological features and those that relate to our intelligence and social needs and inclinations. 'There would be no morality without animality and likewise no morality without capacity-rationality' (p. 23).

Hacker discusses Von Wright's types of goodness in order to illustrate how 'human nature is the source of many kinds of value' (p. 15). Medical goodness is linked to the health and welfare of a living being, its organs and faculties. Technical goodness is the goodness of skills. Instrumental goodness is the goodness of instruments and implements. Beneficial goodness concerns what is good for a being and what does good to a being, and is essentially linked to welfare, prospering and flourishing. Hedonic goodness concerns the good of the pleasant and the pleasing, of the enjoyable and the delightful. Unlike Von Wright, Hacker argues that moral values are not secondary or derivative forms of goodness; but they relate to other forms of goodness and all forms of goodness contribute to the good of a person.

Hence there is nothing mysterious or meta-physical about morality, and our moral values are rooted in our nature and the nature of our world. Our ideas of what is good and what is bad arise from the facts that we are language-using, social beings with rational powers who find ourselves in a world where survival is a constant challenge. Morality is a predictable consequence of the sort of complex life that characterises human beings.

Acknowledging his debt to Aristotle, Hacker highlights how moral values are essentially social—they are about how we conduct ourselves in a group and how we behave towards other people. Hence moral values act as a 'social glue,' uniting a particular group or society around a consensus of right and proper behaviour, which in more developed societies is partially codified into a legal code that sets out proscribed behaviour and associated sanctions.

But Hacker also points out that there are periods in history in which the 'traditional moral order' is called into question. This occurred in ancient Greece as exemplified by the figure of Socrates and some of the Sophists, and again in the Enlightenment starting in Europe in the 17th century, which followed a thousand years of Christian hegemony. The Enlightenment creates new values and ideals—those of questioning and of tolerance. The ideal human changed from the dutiful and faithful Christian 'servant' of the middle-ages to the questioning, open-minded, autonomous individual that we aspire to in the modern age.

Hacker does not develop these ideas further as he has so much else to cover, but there is a lot more to say about how defining features of modernity

such as science and capitalism have moulded our natures and changed what we value about ourselves and our lives. A recent analysis of Marxist ethics, for example, suggests that Enlightenment values inevitably remain limited in a capitalist society based on private property (Blackledge 2012). Several commentators, myself included, have described how neoliberal capitalism has helped to shape our ideas about what it is to lead a 'healthy' and fulfilling life, and how the neoliberal norm is used to define failure as mental disorder (J. Davies 2022; W. Davies 2011; Fisher 2009; Moncrieff 2022).

Hacker's grounding of moral sensibility in the nature of biological life is what makes his account particularly interesting. Other modern philosophers have emphasised the social nature of morality, particularly those associated with virtue ethics, but Hacker shows how the concept of moral goodness relates to more general notions of good and bad that are an integral part of understanding the nature of living things. He highlights how inherent, biologically-based features of human nature and the nature of the world we live in make moral judgements a natural feature of human life. This is not the same as saying that morality follows nature—of justifying survival of the fittest or any other such crude notion. Hacker is certainly not an evolutionary determinist. What he is highlighting is that our moral inclinations arise from our biological nature, and also that our natural inclinations make a moral code a necessity for successful social cooperation. We have a biological predisposition for caring and sympathy with others. At its most basic level this is rooted in the instinct to care for our young, but given our social nature it extends far beyond our offspring or immediate family. We need to live in social groups both for our survival and for the development of our intellectual and physical capacities. But we also have instincts to compete and to survive. Our inclinations need to be balanced and regulated in order for humans to live together successfully.

2. What is the moral good?

In surveying previous moral theories, Hacker exposes false dichotomies. For him, virtue ethics and Kant's deontological ethics are not alternative positions, but two aspects of a 'complex whole' (p. 34). Utilitarianism's opposition of selfishness and altruism, and Kant's contrast between free and

rational behaviour and that driven by desire are both misleading. Pursuing our own inclinations is only selfish if there is a moral demand upon us not to, and acting to satisfy a desire is not causally determined, except in some exceptional cases (see below).

Hacker's fundamental point is that 'Moral goodness is exhibited in one's attitudes towards other people' (p. 37). Serving the interests, welfare and well-being of others or of society as a whole is morally praiseworthy. This is constitutive of moral goodness in almost all cultures. Like Aristotle, Hacker describes how communities embody notions of goodness that determine the characteristics that make a good person; these are the virtues.

What Hacker shares with Kant is a belief that morality presupposes rationality- that is the ability to appraise the world in a sophisticated way and make decisions based on reasons. Our nature as practical animals that react to and act on the world also entails that we care about things. The fact that we are social animals demands that we care about each other. The capacity to care for other people is a presupposition of morality. We have instincts to care about others, such as maternal and paternal instincts to protect and nurture our young, and these are the primitive roots of our moral values, but we are also taught and learn to care.

A key development in our moral outlook that Hacker dates to the Enlightenment is what he calls 'formal respect;' that is the idea that every human being deserves respect and dignity (non humiliation) by virtue of being a human being—not just a member of a particular community or group. Although we now take this idea for granted, and it is enshrined in the concept of 'human rights' (which Hacker sees as a related but narrower concept due to its legal rather than philosophical foundations), formal respect is not only a relatively recent idea, it is also fragile. It was most obviously rejected by Nazism and apartheid South Africa but even today, discrimination against people on the basis of race, sex and sexuality across the globe can be seen as subverting the principle of 'formal respect'.

Hacker credits Kant for establishing the principle of formal respect and although he disagrees with how Kant sets out the categorical imperative, he seems to me to align with Kant in basing the principle of respect on our shared rationality. The fact that human beings are rational agents who make free choices about how they act and are therefore responsible for what

they do makes every human being worthy of respect. Recognising the freedom of others entails respecting their autonomy.

This raises the question of how we should treat people who lack fully developed rational capacities, including small children, people with intellectual disabilities and those who have suffered brain damage or disease. Formal respect is not something that applies to individuals according to their capacities, it is a principle that applies to everyone by virtue of their membership of the human race. Yet this entails that what 'formal respect' means in practice can vary, depending on the capacities of the individual. We may respect the right to life and freedom from cruelty for all human beings, but children are not generally allowed the same freedom as adults, and those with limited intellectual capacities may not be granted the same rights and privileges as others.

Hacker points out the apparent contradiction that the ancient philosophers recognised the rationality of man but did not make the leap to formal respect because they lived in a slave owning society. The idea of formal respect was incompatible with the social structure of the ancient world, and there was no significant impetus for changing this structure.

Here, again, Hacker highlights how social and economic conditions influence our moral thought, which is reminiscent of Hegel and Marx (Hegel 1976; Marx 1993). It is only with the rise of commercial and later industrial capitalism that we get a fully-fledged notion of the individual (Blackledge 2012). There are many harbingers of individualism. The Reformation is often thought of as a significant step towards individualism, though whether it reflects the social and economic conditions of emerging capitalism in a Marxist sense, or creates them as Weber claimed (Weber 1958), is a matter of debate. Nevertheless, Enlightenment thought and the principle of formal respect that Kant articulates reflect the new social relations introduced by the fall of feudalism and the rise of capitalism, and particularly the emergence of a working class with the power to demand recognition.

3. Virtues

Hacker argues that the characteristics that are considered good or virtuous have changed little throughout human history, and that we tend to

overlook some of the constant features of our moral outlook until particular historical junctures bring them into view. Such is the 'golden rule' of treating others as you would be treated yourself. Hacker traces the articulation of this principle to the House of Hillel in the 1st century BC, although Christianity must be credited with popularising and disseminating it. The implication is that this principle of according other people the kindness you would wish for yourself is a deeply ingrained one that transcends historical epochs and geographical and cultural boundaries, even if it is only made explicit in certain conditions.

The continuity of valued characteristics such as kindness, generosity and trustworthiness that are oriented to the interests of others is readily apparent. However, one can stress the constancy of the virtues or their transmutability, and both positions seem correct and important. Although Hacker's examples of how virtues change their importance is meant to illustrate the relative triviality of these changes against the constancy of the backdrop, again his analysis illuminates how our material conditions shape our values. Nietzsche complained about how Christianity had ousted the values of bravery, honour and courage that characterised the masters of warlike societies and made the slave values of weakness and submission into virtues. Hacker adds that the modern welfare state has made charity less important, improved medicine and pain relief have rendered fortitude and endurance less significant, and that the availability of effective birth control means chastity (a virtue that has mainly applied only to women, of course) is no longer highly prized. Although lust, as Hacker points out, is still technically a vice, it plays little role in contemporary moral tales. The Christian values of faith (which as Hacker points out is not a virtue if one doesn't believe there is anything to be faithful to) and pity, do not match modern sensibilities, although mercy would surely still qualify.

Hacker has little time for Nietzsche although he welcomes his rejection of a religious foundation to morality and his challenging of medieval Christian 'martyrology' and 19th century hypocrisies. Hacker's answer to the spectre of relativism raised by Nietzsche seems too brief, however, given how compelling this view has become and remains, and given that, as Hacker freely admits, many currently existing human societies do not share the Enlightenment values that Hacker advances. His argument is that these

values are the most conducive to the Good of Man and, if adopted by a society, give everyone the best chance to flourish- that is to fulfil their potential and lead a meaningful life. Moreover, living according to these values enables each of us to find our 'own soul' (p. 64) as Hacker puts it, so that it is not just that they are good for other people and the community in general, they are good for each of us in the sense that they enable us to live a good and meaningful life.

Again, this is similar to Marxist arguments that the highest moral values are those that enable the realization of each and every individual's 'species being,' although Marxists would argue that this is only possible in a socialist system where significant material inequality is abolished (Blackledge 2012). The problem with this position is that much of the world is still not convinced of the superiority of the Enlightenment view, and it is not clear how it can be persuaded. Western values have increasingly become equated with colonialism and the oppression of indigenous cultures abroad, and at home are under attack from those who feel disorientated by freedom and desire a return to more traditional values (witness the recent overturning of the right to abortion in the United States). Hacker is aware of this and concerned about the fragility of the Enlightenment project. As Marxists suggest, if morality is 'a constantly contested product of historical conditions', then only political action can change moral outlooks, but increasingly there seems no guarantee that the progressive side will triumph (Moncrieff 2014) (p. 63).

4. Freedom

For Hacker, the fact that our actions are free is inherent in the idea of human action. Without it, we are dealing not with actions, but with reflex movements. As Hacker puts it: 'we are free agents tracing a spacio-temporal path through the world in accordance with our inclinations, preferences, choices, intentions and decisions and in pursuit of our goals, subject to chance and fortune' (p. 161).

We are not bundles of particles whose trajectory is determined at the atomic or subatomic level. All complex biological beings are highly structured and subject to 'top-down' control. Any form of reductionism that

seeks to explain the behaviour of living beings in terms of their atomic make-up, their chemical composition or the structure of their nervous system, for example, ignores all explanations that are distinctive of biological creatures. Reductionist explanations cannot account for the propensities of living things because these can only be understood at the level of the organism (notwithstanding Dawkins attempts to ascribe them to genes). Neither can they even account for the workings of bodily organs, which also have to be understood in a functional sense, as serving the organism as a whole. And most importantly, reductionist explanations cannot account for the individual actions of living creatures, because these cannot be predicted in a mechanical manner. As biologist Steven Rose pointed out, the future of living beings is inherently indeterminate (Rose 1997).

Purposiveness and rationality render actions meaningful. In other words, we can understand actions in terms of how they help or hinder an individual's aims and intentions. We explain actions in this way by giving reasons, which answer the question as to why someone did what they did. Nothing that has meaning can be explained by a mechanistic account of how it came to be. 'An essentially mechanical world would be an essentially meaningless world' (p. 177) as Neitzsche put it in the Gay Science (quoted by Hacker).

It is curious that determinism exerts so much appeal to philosophers and scientists, since, as Hacker points out, no one behaves as if it were true—indeed, it is doubtful that anyone even could act as if it were true. We respond to each other and to animals as beings that make free choices, and it is difficult to imagine how we could behave otherwise. Maybe the appeal of determinism reflects the existential crisis produced by the decline of religion. As Hacker pointed out in the first volume of his series on human nature, The Categorial Framework, religious design brought teleology in general into disrepute, and for many thinkers, ideas about purpose conjure only religious meanings (Hacker 2010). Perhaps, it takes an outlook that is not reacting against a religious worldview to see purpose and meaning in other ways.

Hacker helpfully distinguishes determinism in its modern form from the much older idea of Fate and fatalism, common to many ancient cultures. Fatalism is not the idea that our actions are pre- determined by antecedent conditions, such as the state of our brains, but that life is subject to chance,

and that our abilities to influence and control our environment are necessarily limited. The ancient idea of Fate is a personification of all that is beyond the sway of human beings. Earlier cultures recognised that it is necessary to accept this fact to face life with equanimity 'for to rail against fortune is futile, and to resent it is to undermine one's ability to live wisely within its constraints' (p. 166). But in the modern world where we exert so much more control over many aspects of our environment, we find this difficult. We may, for example, have unrealistic expectations about medicine's ability to conquer and cure all forms of disease. We certainly have unrealistic aspirations that we can eliminate sadness, depression and less pleasant emotions in general, a situation that is suggested to be partly responsible for our modern epidemic of mental health problems (Timimi 2021).

5. Neuroscientific determinism

Hacker is one of the most articulate critics of neuroscientific determinism and devotes a chapter to the subject and its ramifications, based on his previous books and papers on the subject (Bennett & Hacker 2003; Nachev & Hacker 2014).

Neuroscientific determinism involves the idea that we can predict certain forms of behaviour from the state of the brain, but for Hacker the idea that we might be able to read our thoughts and behaviour off the brain is nonsensical. What we say and do can only be made sense of in the context of the human world, it cannot be explained by talking about brain events or states. We might sensibly say that there are correlations between certain neural activity and muscular contractions, but not with 'agential actions as opposed to mere movements, let along moves in a language game of a human community at a given stage in human history' (p. 181). The fact that we can discover certain functions of the brain through studying the deficits produced by various diseases has enticed us into the belief that we can understand normal human behaviour (that is behaviour that is not driven by a disease process) through the workings of the normal brain, but as Hacker suggests, such beliefs are rooted more in science fiction than science.

There are many explanations for our actions and behaviour, Hacker reminds us (he covered this ground in volume 1, The Categorial Framework). Mechanistic, causal explanations are one very specific type of explanation that apply in certain narrow circumstances. Even then, they are rarely related to neurological factors, and more usually involve environmental conditions, such as when one slips on the apocryphal banana skin. There is a small collection of behaviours which we attribute to neurological diseases in a classical causal sense. Yet most human behaviour requires explanation in terms of reasons and motives. If someone is writing a letter, a neurological explanation can, at most, explain the nature of the movements involved. It will not explain the nature of the activity, nor why it was undertaken.

Explanation, as Hacker reminds us, involves making something understandable. Neurological descriptions do not make human behaviour understandable except in a few very specific situations where a brain disease or injury has caused an alteration in someone's behaviour.

Hacker reiterates his previous response to the famous Libet experiment that appeared to show that neural activity precedes the decision to act (Nachev & Hacker 2014). His criticisms derive from his understanding of human capacities as being dependent on the good functioning of the brain, but not inherent in the brain. In Hackers view, which seems compelling, the action potential that can be detected in the brain that precedes the conscious decision (the latter indicated by the action of pressing a button), indicates that a state of neural readiness is necessary for action to take place. Neural activity makes it possible for us to act, but does not make it necessary.

The misinterpretation of the Libet experiment as indicating that our decisions are determined by our brains is due to the common mistake of seeing intentions and thoughts as events in the brain that act as mechanical causes of action. But they are neither of these. Decisions are not always concurrent with the actions they relate to—one can decide in advance that one wants to do something, and one can decide that one does not want to do something, but it makes no sense to think of the causes of non-actions.

Mental capacities are properties of persons not brains, although they depend, of course, upon having a well-functioning brain. Decisions and intentions are made by persons, for example, not brains. Knowing and

remembering are also capacities of persons. Knowledge is not stored in the brain as memories. Instead, memory consists of knowing now things one previously came to know or apprehend. It is the power to retain previous knowledge and abilities.

Character traits are also not located in the brain. They are 'tendencies and pronenesses' of persons (191), that manifest themselves in repeated patterns of behaviour whose interpretation crucially depends on its particular context. Being a shy person is characterised by relative timidity of behaviour in certain social situations compared to other people in the same situation. But in other circumstances, such timidity might be what is normally expected, and hence the same behaviour would not count as shyness. The anxiety that someone might feel in a social situation is not shyness either, although it is part of it—but if it is not manifested in behaviour in any way we would not normally consider the person to be shy.

6. Responsibility

By virtue of our capacity to reason, to reflect on our circumstances and weigh up our options, we are responsible for what we do, and we are unique among animals in this respect. However, Hacker makes the important point that 'the concept of responsibility is neither clear-cut nor distinct'. He also highlights how the concept of responsibility rests upon the notion of a human being with normal capacities for rational deliberation and action. People can be held responsible for their actions in so far as they know what they are doing, understand their situation, are capable of reflection and forming intentions and can exert control over their actions—but these criteria are not necessarily straight-forward.

Various circumstances can interfere with the capacities we need to be considered responsible for our actions. In the United Kingdom in 1843, the McNaughton rules excluded someone who had a 'defect of reason, as not to know the nature and quality of the act he was doing, or, if he did know it, that he did not know that what he was doing was wrong' from criminal liability for murder – sometimes referred to as the insanity defence. The issue was the individual's ability to reason and to have knowledge of his or her situation specifically, and of moral norms in general. In other countries,

Hacker tells us, the same sort of legislation was broader and encompassed those who had a 'defect of will' or an inability to control their emotions as well as a defect of reasoning power. Although England introduced something similar in the Homicide Act of 1957, it appears that different legal jurisdictions interpret the criteria for responsibility differently.

At issue is whether a loss of self-control or generally poor self-control can be considered to excuse responsibility. This is relative to circumstance, and as Hacker suggests most of us would excuse the victim of torture who spills the beans on his comrades, even if we might admire the individual who managed not to. But what of the man who kills his wife in a fit of rage or the women who kills her husband after years of abuse and humiliation? Do we excuse the alcoholic who leaves his family destitute? These are less clear-cut.

Hacker wants to include defects of will in those situations in which we excuse people from responsibility for their actions. He notes that fellow philosopher Anthony Kenny objects to this on the grounds that there is no way of distinguishing an irresistible impulse from an unresisted impulse. In other words, we cannot know how much someone may have tried to resist an impulse that they eventually give in to, or indeed whether they tried to resist it at all. Moreover, our judgements are always relative. We generally excuse people if most people fail to control their urges in similar situations (e.g. submitting to torture) but not if we think most reasonable people would resist (e.g. rape , although our attitudes are culturally sensitive).

Hacker feels that situations in which the will is impaired—either through the extreme nature of the circumstances or through an addiction such as alcoholism – are only partly voluntary. The nature of the situation or the addictive impulse over-rides the ability to make fully autonomous, free choices. He believes this should be recognised legally – that impairment of the will should excuse legal responsibility and that this should apply to cases of addiction.

I do not fully agree with Hacker here. I do agree that failing to control one's impulses and emotions is a common phenomenon. Our ability to do this is not only determined by immediate circumstance, it is also shaped by our personal history, including the luck of our birth and upbringing. And

for this reason we should extend our understanding and have sympathy for many of those who find it difficult to resist certain forms of behaviour.

We do not always need to punish people for their misbehaviour, but should we excuse them of responsibility? Bringing one's behaviour into line with social norms and expectations is a moral obligation, as Hacker points out. He also points out how mental states and inclinations are not mechanical causes of behaviour. Indeed, he admits that acting on impulse is 'not to be caused to act by a mysterious mental cause denominated an impulse' or, we could add, 'addiction.' Therefore, the behaviour remains the behaviour of the individual—that is behaviour that is freely initiated by them. Of course, all our actions are limited by circumstance, and conditioned by the person we are and have become, with all the developmental history that goes into making each one of us who we are. But absent a neurological condition, the behaviour remains the behaviour of the individual.

This also applies to 'defects of reason' when these occur in the context of a mental disorder, such as schizophrenia, rather than a neurological one. When people lose their ability to reason by becoming immersed in a fantasy world, or by withdrawing from the shared, social world what they do is still attributable to them.

We should also look at the other side of the coin. If we excuse someone of responsibility for their actions, we also deny that those actions are fully autonomous. This may not matter if the action is a one off, but if it is something someone does recurrently, part of a pattern of behaviour, this becomes a denial of the individual's autonomy per se. It is tantamount to saying that this person does not count as a full person, and that they need to be treated as a child or as someone who is mentally impaired. When we do this in today's society, it entails the right to do various things to people against their will—such as incarcerating them or forcing them to take mind-changing drugs. It has been used as a reason not to extend people the right of 'formal respect,' as the Nazis did when they exterminated the mentally ill, and many western countries did with the sterilisation of the mentally ill in the early 20th century. The cancelling of responsibility does not come for free.

Where I agree with Hacker is that we should extend sympathy to people on the basis that struggling to control one's impulses and emotions is

a universal human experience, and that the circumstances of people's lives, coupled with natural variation in tastes and inclinations, will make this more difficult for some than for others. Being held responsible need not be linked to punishment, and as we already do, extenuating circumstances can be taken into account when considering how to 'dispose' of people who have done wrong.

7. Conclusion

Hackers anthropological account of morality peels away the mystery and embeds moral values in the nature of biological life in general, and the features of human life in particular. This does not do away with the problems posed by relativism, but it does provide a sound starting point that highlights the important constants in human values, against which we can judge the way these have also been shaped by particular social conditions.

Moral values are inherently social; they involve ways that we behave towards other individuals and our community in general. Hacker's analysis, like others', highlights how moral values change in response to changes in the organisation and economic basis of society, and how some periods of history witness epochal changes in the nature of these values. The Enlightenment represents the most recent such change, ushering in new values of tolerance and respect for all human beings for the fact that they are human beings, regardless of race, sex, status or creed. Modern Enlightenment morality can be judged to be superior to other moral codes in that it better enables the flourishing of all human beings, and through this the flourishing of the human community as a whole.

Hacker's book underlines how morality also necessarily implies that human beings are free—free to make choices within the restrictions of history and circumstance—for good or for ill. Determinism and morality are conceptually contradictory and the fact that we have moral propensities and moral language is one reason among many to conclude that determinism cannot be correct. The richness of human life, including our inclination to distinguish good from bad and right from wrong cannot be reduced to brain activity, although our biological nature, including our large brains, are what

make this possible. 'Men (and women) make their own history', as somebody once said (Marx, 1852).

References

Bennett, M. R., & Hacker, P. M. S. 2003. *Philosophical Foundations of Neuroscience*. Oxford: Blackwell Publishing.

Blackledge, Paul. 2012. *Marxism and Ethics: Freedom, Desire, and Revolution*. New York: Suny Press.

Davies, James. 2022. *Sedated: How Modern Capitalism Created our Mental Health Crisis*. London: Atlantic Books.

Davies, William. 2011. "The Political Economy of Unhappiness". *New Left Review*, 71(Sept/Oct). Retrieved from https://newleftreview.org/issues/ii71/articles/william-davies-the-political-economy-of- unhappiness

Fisher, Mark. 2009. *Capitalist Realism: Is There No Alternative?*. London: Zero Books.

Hacker, P. M. S. (2010). *Human Nature: The Categorial Framework*. Oxford: Wiley-Blackwell.

Hegel, G. W. F. 1976. *Phenomenology of Spirit* (A. V. Miller, Trans.). Oxford: Oxford University Press.

Marx, Karl. 1852. *The Eighteenth Brumaire of Louis Napoleon*. New York: Die Revolution.

Marx, Karl. 1993. "A Contribution to the Critique of Political Economy". *Preface*: Marx.org.

Moncrieff, Joanna. 2014. "FDR: Review of Marxism and Ethics: Freedom, Desire and Revolution by Paul Blackledge". *Radical Philosophy* 185(May/June): 62–3. Retrieved from https://www.radicalphilosophyarchive.com/issue-files/rp185_reviews.pdf

Moncrieff, Joanna. 2022. "The Political Economy of the Mental Health System: A Marxist Analysis". *Front Sociol* (6): 771875. doi:10.3389/fsoc.2021.771875

Nachev, Parashkev & Hacker, Peter. 2014. "The Neural Antecedents to Voluntary Action: A Conceptual Analysis". *Cogn Neurosci* 5 (3-4): 193–208. doi:10.1080/17588928.2014.934215

Rose, Steven. 1997. *Lifelines: Biology, Freedom, Determinism*. London: Allan Lane.

Timimi, Sami. 2021. "Insane Medicine: How the Mental Health Industry Creates Damaging Treatment Traps and How you can Escape Them: Independently". Published online

Weber, Max. 1958. *The Protestant Ethic and the Spirit of Capitalismm and Other Writings* (T. Parsons, Trans.). New York: Scribner/Simon & Schuster.