Contents

Research Articles

Organon F 30 (4) 2023: 312–322
https://doi.org/10.31577/orgf.2023.30401
ISSN 2585-7150 (online)
ISSN 1335-0668 (print)

RESEARCH ARTICLE

The Simplest Solution to the Deepest Paradox of Deontic Logic

Jeremiah Joven Joaquin*

Received: 10 May 2023 / Revised: 20 October 2023/ Accepted: 22 October 2023

Abstract: Since its inception in 1984, several 'simple' solutions have been proposed to answer the deepest paradox of deontic logic. In this paper, I present the simplest one yet: the deepest paradox is simply ill-formulated.

Keywords: Deontic logic; paradox of gentle murder; the deepest paradox of deontic logic; imperatives; indicatives.

1. Introduction

In a talk presented in 1985, Hector-Neri Castañeda dubbed James William Forrester's gentle murder paradox as 'the deepest paradox of deontic logic' (Meyer, 1987; Goble, 1991). Forrester's paradox is a variant of the good Samaritan paradox that Arthur Prior (1958) raised a few decades before.

Many philosophers agree that Prior's paradox has a simple solution: it confuses the scope of deontic operator, $\mathbb{O}$. This solution implies that the scope of $\mathbb{O}$ are actions, not propositions or imperatives (see, e.g., Nozick

* De La Salle University
https://orcid.org/0000-0002-8621-6413
Department of Philosophy, De La Salle University, 2401 Taft Avenue, Mal- ate, Manila, Philippines 0922
jeremiah.joaquin@dlsu.edu.ph

and Routley (1962); Åqvist (1967); Sinnott-Armstrong (1985)). For some, e.g., Sinnott-Armstrong (1985), the scope-confusion solution also applies to the gentle murder paradox. For others, the simple solution lies elsewhere, e.g., in dynamic logic (Meyer, 1987), in world-semantics (Feldman, 1990), in an optimal satisfaction model (Kowalski and Satoh, 2018), and so on. As of now, there is no consensus as to what the correct 'simple' solution to Forrester's paradox is.

In this paper, I argue that the simplest solution to the gentle murder paradox is to show that it is simply ill-formulated, and once this is pointed out, the paradox disappears. But before getting into this, let us first rehearse Forrester's paradox.

2. The deepest paradox

Forrester (1984) starts with two legal assumptions about murder:

(A1) All kinds of murder are forbidden.[1]

(A2) Murdering someone violently is worse than murdering someone gently.

and three standard principles of deontic logic:

(D1) $\mathbb{O}(p \supset q) \supset (p \supset \mathbb{O}q)$

(D2) $(p \supset q) \Rightarrow (\mathbb{O}p \supset \mathbb{O}q)$

(D3) $\mathbb{O}p \supset \sim\mathbb{O}\sim p$

(Note: we use the basics of deontic logic in formulating (D1)-(D3). p and q are well-formed formulas. '$\sim$' represents negation and '$\supset$' the material conditional. '$\mathbb{O}$' is the deontic operator, 'It is obligatory...', and '$\Rightarrow$' indicates entailment.)

(D3) is the familiar D-axiom of certain deontic logics, viz., 'Whatever is obligatory is permissible', where 'It is permissible that p' is defined as 'It is not obligatory that not-p. (D1) is a principle about conditional obligations,

[1] Alternatively, (A1) can also be formulated as 'It is obligatory that no one does any kind of murder'.

viz., 'If q is obligatory given p, then if p is the case, q is obligatory. Finally, (D2) is the controversial inferential rule that from 'If p, then q' it follows that 'If p is obligatory, then q is obligatory'.

Given this, the paradox proceeds as follows. From (A1) and (A2), we respectively have:

(1) It is obligatory that Smith does not murder Jones.

(2) It is obligatory that if Smith murders Jones, Smith murders Jones gently.

The thought for (1) is that if murder is forbidden, then it is obligatory that no one does it. On the other hand, the thought for (2) is that if murder is unavoidable, then it must be done gently (rather than violently).

From (2) and (D1), via modus ponens, we derive:

(3) If Smith murders Jones, it is obligatory that Smith murders Jones gently.

Forrester then makes a further assumption. Suppose that, in fact,

(4) Smith murders Jones.

From this assumption and (3), via modus ponens, we derive:

(5) It is obligatory that Smith murders Jones gently.

But any form of murder (gentle or otherwise) *is* murder, so, via logic:

(6) If Smith murders Jones gently, then Smith murders Jones.

Thus, from (5) and (6), using (D2) and modus ponens, we derive:

(7) It is obligatory that Smith murders Jones.

Finally, from (7) and (1), via (D3) and adjunction, we derive the contradiction:

(8) It is and it is not obligatory that Smith murders Jones.

The gentle murder paradox differs from other versions of the good Samaritan paradox because of the adverbial steps (2) and (5) (and all the derivations from them) and (D1) that 'blocks a scope solution to the paradox' (Forrester, 1984, 195–96).

3. The simplest solution

I agree with Forrester that the standard scope-confusion solution to the gentle murder paradox does not work. However, such a solution could be refined to show that the main problem with this paradox is its formulation.

Forrester's paradox (and other deontic paradoxes of the same ilk) assumes that the scope of $\mathbb{O}$ is actions or indicatives of actions of the form, 'Some subject, *S*, does some action, *A*'. I think this is a mistake. At least in deontic paradoxes that talk about what agents are obligated to do, the scope of $\mathbb{O}$ is not actions or indicative of actions but imperatives or prescriptions of actions.

Indicatives of actions are descriptions of someone doing (has done or will do) a particular action. For example, 'Amanda is cleaning her room' describes the fact that Amanda is cleaning her room. If the scope of $\mathbb{O}$ is indicatives of this sort, then '$\mathbb{O}$(Amanda cleans her room)' means 'It is obligatory that Amanda cleans her room'. One reading of this implies that the fact that Amanda is cleaning her room is obligatory. However, this seems wrong since it means that some fact is obligated to do something. But how can this be? Facts do not seem to have an obligatory quality. Facts happen, may happen, necessarily happen, or cannot possibly happen. But they cannot be obligated to happen. Nor are they the sort of things that could be obligated to do something.

Of course, there is a case where the scope of $\mathbb{O}$ might be fact. For example, consider the subjunctive function of ought in 'The world *ought to have been* without pain', where the corresponding fact is that the world *is* with pain. This sentence does not suggest any course of action that someone must take. It only indicates a value judgement that someone may have. Such a judgement may have the form, 'The world would have been *better* if there were no pain in it' (Castañeda, 1975, 46). However, this is arguably not the sort of 'ought' at stake in Forrester's gentle murder paradox, which talks about what an agent, viz., Smith, is obligated to do.

Since the sort of paradox that we are concerned with in this paper is a paradox about what agents ought to do, the scope of $\mathbb{O}$ needs to be the sort of thing that *drives* agents to do a particular action. That is, it is the sort of thing that might force agents to do something. Obviously, different

factors motivate people to do something. But what we are concerned with here is a sort of thing on which 𝕆 might be attached. And I submit that this sort of thing is imperatives.

When we issue imperatives or prescriptions of the form '*S*, do *A*' or, more simply, 'Do *A*' (for some subject, *S*, and some action, *A*), we are mandating some agent to do a particular action. For example, if I were to ask Amanda to clean her room, I would do it with an imperative, 'Amanda, clean your room'. Or, if I am gentle, I will make a request, 'Amanda, please clean your room'. But in either case, the form of my utterance is not a statement of fact but a mandate whose content is a prescribed action, which is either complied with or not.[2] Given this, it follows that '𝕆(Amanda cleans her room)' does not mean that some fact, the fact that Amanda cleans her room, is obligatory; rather, it means that I am obligating Amanda to clean her room. I am saying, 'It is obligatory *for* Amanda *to* clean her room'.

Now, one may ask about conditional obligations of the form, 'If *p*, then one must do *q*' (where 'must' indicates an obligation). How should we understand them? That is, what does '𝕆(If *p*, do *q*)' mean if the scope of 𝕆 are imperatives? The short answer here is that '𝕆(If *p*, do *q*)' means 'It is an obligation to do *q* on the condition that *p* occurs'. For example, suppose that *p* is the indicative, 'The stoplight is red', and *q* is the imperative, 'Stop'. Then '𝕆(If *p*, do *q*)' means 'It is obligatory for someone to stop given that the stoplight is red'.

Given this, we now have a reason to accept an apt reformulation of Forrester's (D1) and (D3) and to reject his premises (1) and (2) and his formulation of (D2). On the one hand, given that the scope of 𝕆 is imperatives (and not indicative of actions), (D1) must be understood as being about conditional obligations, with imperatives as their constituents. For instance, given our traffic rules, 'It is obligatory (for someone) to stop on the condition that the stoplight is red' entails 'If the stoplight is red, then

[2] An interesting issue about imperatives is their truth-aptness. That is, whether they have truth conditions. A related problem is that if such imperatives do not have truth conditions, how could they be constituents of valid (deductive) arguments? Several proposals have been made regarding these issues (see, e.g., Parsons (2013) and Joaquin (2022)). However, we will not touch on them here.

it is obligatory (for someone) to stop'. Likewise, (D3) must be understood along the same lines. If it is obligatory (for someone) to stop (given some condition), it is permissible (for that someone) to stop (given the very same condition).

On the other hand, Forrester's premises (1) and (2) suffer a scope confusion given that the scope of the ought operator is imperatives and not indicatives of actions. Thus, we may reject them outright. Finally, we can reject (D2) since there are counterexamples to it if its instances are indicatives (and not imperatives). For example, 'If Amanda's room is dirty, then Amanda cleans it' does not entail 'If it is obligatory that Amanda's room is dirty, then it is obligatory that Amanda cleans it' simply because the antecedent of the latter conditional does not make sense. As was argued above, it does not make sense to say of a given fact that it is obligatory. In particular, how can the fact that Amanda's room is dirty be obligatory?[3]

We now have the simplest solution to the deepest paradox of deontic logic (and, more generally, to deontic paradoxes concerned with what agents are obligated to do). If the scope of $\mathbb{O}$ (or any deontic concept) is imperatives (and not indicatives of actions), then the paradox is simply ill-formulated.

4. A simple riposte?

Now, some theorists may remain unconvinced. They may try to reformulate Forrester's paradox in terms of imperatives and not indicatives of actions. In particular, they may reformulate the controversial inferential rule (D2) as follows:

(D2*) $(p! \supset q!) \Rightarrow (\mathbb{O}p! \supset q!$ (where $p!$ and $q!$ are imperatives)

Given this, they may have a revised gentle murder paradox as follows:

[1] It is obligatory for Smith does not to murder Jones. (From (A1))

[3] Forrester acknowledges that (D2) might be the 'rotten apple [in] the entire barrel of standard deontic logic' (Forrester, 1984, 197). However, he did not develop this point in his paper.

[2] It is obligatory for Smith to murder Jones gently if Smith murders Jones. (From (A2))

[3] If Smith murders Jones, it is obligatory that Smith murders Jones gently. (From [2] via (D1))

[4] Smith murders Jones. (Assumption)

[5] It is obligatory for Smith to murder Jones gently. (From [3], [4] via modus ponens)

[6] If Smith murders Jones gently, then Smith murders Jones. (Fact)

[7] It is obligatory for Smith to murder Jones. (From [5] and [6] via (D2*) and modus ponens)

[8] It is and it is not obligatory for Smith to murder Jones. (From [7] and [1] via (D3) and adjunction)

However, formulating the paradox this way makes explicit what is wrong with Forrester's gentle murder paradox, viz., the faulty reasoning from [5] to [7] via (D2*) in the revised paradox, and the corresponding faulty reasoning from (5) to (7) via (D2) in the original paradox. On the one hand, even if we grant [5] and (D2*) in the revised paradox, the crucial inferential step from the conditional with factual contents (indicatives) in [6] to the deontic *statement* in [7] does not hold since there is no inferential rule or deontic principle that warrants the entailment from facts to what an agent is obligated to do.[4] (This is reminiscent of Hume's no-ought-from-is principle.[5]) On the other hand, the crucial inferential step from (5) to (7) in the original gentle murder paradox does not work since, as was discussed above, (D2) does not hold if its instances are indicatives of actions, not imperatives.

[4] Saying that [7] is deontic *statement* implies, more generally, that despite having an imperative constituent, sentences of the form, 'It is obligatory for S to do A', are truth-apt; they can be true or false.

[5] For a discussion of Hume's principle, see Pigden (2010).

5. An homage to Castañeda

The general idea proposed in this paper is nothing new. It follows the deontic theory put forward by Castañeda (1960, 1968, 1970, 1975).[6] In particular, my proposal follows his distinction between *propositions* and *prescriptions* or *practitions*, on the one hand, and his distinction between an ought-to-be and an ought-to-do deontic judgement (statement), on the other hand.[7]

For Castañeda, propositions are the bearers of truth-values. They are typically expressed in indicative sentences. They are the objects of belief, knowledge, and other so-called propositional attitudes. And they figure in entailments. On the other hand, practitions, which include intentions, prescriptions, requests, mandates, commands, and imperatives, are not truth-apt *per se* but could nonetheless figure in entailments (Castañeda, 1975, Ch. 4). For example, the conjunctive imperative, 'Amanda, go and clean your room', entails the simple imperative, 'Amanda, go to your room' or 'Amanda, clean your room'.

A reader knowledgeable of Castañeda's theory might notice that his idea of practitions informs how I used imperatives in this paper. However, my usage is simpler than Castañeda's. His theory has a complex metaphysics involving noematic structures that make propositions semantically akin to practitions (Castañeda, 1975, 7). I do not share (nor do I need) such metaphysics. My distinction between indicatives and imperatives, while purely at the linguistic level, is enough to address the deepest paradox of deontic logic.

What I do share with Castañeda's theory is the thought that deontic statements about what agents are obligated to do imply that the scope of $\mathbb{O}$ is imperatives (practitions) and not indicatives of actions. This is an essential component of this paper's proposal. And Castañeda distinction

[6] In fact, Forrester (1984, 193) acknowledges Castañeda's persuasion as the cause for his 'identification of the likely culprit' of the faulty step (viz., (D2)) in his paradox.

[7] It is instructive to note that Castañeda (1968) used these distinctions to address Prior's good Samaritan paradox – a paradox that the gentle paradox was based on.

between an ought-to-be and an ought-to-do deontic statement is important in this respect.

Recall that in §3, I discussed the idea that the scope of $\mathbb{O}$ might be a fact, and I used the subjunctive deontic statement, 'The world ought to have been without pain' as an example. This is Castañeda's example, and it illustrates an ought-to-be deontic statement – an evaluative statement that does not suggest any course of action that someone must take. In contrast, ought-to-do deontic statements are statements that do suggest (mandate, command, or request) a course of action that an agent must take (Castañeda, 1975, 46).[8] This distinction is important since it is the latter sort of deontic statement that the gentle paradox and other deontic paradoxes of the same ilk are about.

However, like Castañeda, I leave 'undecided whether the Ought-to-be is reducible to the Ought-to-do or vice versa'; notwithstanding that agentless deontic statements, e.g., 'Every car ought to have a licence' can be reduced to an agential, ought-to-do deontic statement like, 'Every car *owner* ought to have their car licensed'. Moreover, like Castañeda, I emphasise the crucial role of agency in thinking about ought-to-do deontic statements and how such thinking might resolve some issues in deontic logic (ibid.).

6. Conclusion

In summary, I argued that the simplest solution to the deepest paradox of deontic logic is to show that the scope of the deontic operator involved in ought-to-do deontic statements is not actions or indicatives of actions but imperatives or practitions. Any deontic paradox that confuses this can easily be (dis)solved.

Disclosure Statement

The author(s) reported no potential competing interest.

[8] Von Wright (1999) makes a similar distinction between *Sein-Sollen* (ought to be) and *Ton-Sollen* judgements.

Acknowledgements

My thanks to Hazel T. Biana, Ben Blumson, Brian Garrett, Alan Hajek, Stephen Hetherington, Graham Priest, Raymond R. Tan, and the anonymous referee of this journal for helpful comments and suggestions that greatly improved the paper.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

References

Åqvist, Lennart. 1967. "Good Samaritans, Contrary-to-duty Imperatives, and Epistemic Obligations." *Noûs*, 1(4): 361–79. https://doi.org/10.2307/2214624%E2%80%A2

Castañeda, Hector. N. 1960. "Imperative Reasonings." *Philosophy and Phenomenological Research*, 21(1): 21–49. https://doi.org/10.2307/2104787%E2%80%A2

Castañeda, Hector N. 1968. "Actions, Imperatives, and Obligations." *Proceedings of the Aristotelian Society*, 68(1): 25–48. https://doi.org/10.1093/aristotelian/68.1.25

Castañeda, Hector N. 1970. "On the Semantics of the Ought-to-do." *Synthese*, 21(3–4): 449-68. https://doi.org/10.1007/978-94-010-2557-7_21

Castañeda, Hector N. 1975. *Thinking and Doing: The Philosophical Foundations of Institutions*. Boston: D. Reidel.

Feldman, Fred. 1990. "A Simpler Solution to the Paradoxes of Deontic Logic." *Philosophical Perspectives*, 4: 309–41. https://doi.org/10.2307/2214197%E2%80%A2

Forrester, James W. 1984. "Gentle Murder, or the Adverbial Samaritan." *Journal of Philosophy*, 81(4): 193-97. https://doi.org/10.2307/2026120%E2%80%A2

Goble, Lou. 1991. "Murder Most Gentle: The Paradox Deepens." *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, 64(2): 217–27. https://doi.org/10.1007/BF00404830

Joaquin, J. Joaquin. 2022. "Infectious and Transparent Emotivism." *Journal of Applied Non-Classical Logics*, 32(1): 1–10. https://doi.org/10.1080/11663081.2021.2016242

Kowalski, Robert. and Satoh, Ken. 2018. "Obligation as Optimal Goal Satisfaction." *Journal of Philosophical Logic*, 47(4): 579–609. https://doi.org/10.1007/s10992-017-9440-3

Meyer, J.-J. 1987. "A Simple Solution to the 'Deepest' Paradox in Deontic Logic." *Logique et Analyse*, 30(117/118): 81–90.

Nozick, Robert. and Routley, Richard. 1962. "Escaping the Good Samaritan Paradox." *Mind*, 71(283): 377–82. https://doi.org/10.1093/mind/LXXI.283.377

Parsons, Josh. 2013. "Command and Consequence." *Philosophical Studies*, 164(1): 61–92. https://doi.org/10.1007/s11098-013-0094-x

Pigden, Charles R. 2010. *Hume on Is and Ought*. New York: Palgrave Macmillan.

Prior, A. N. 1958. Escapism: The Logical Basis of Ethics. In *Essays in Moral Philosophy*, edited by A. I. Melden, 135-146. Seattle: University of Washington Press.

Sinnott-Armstrong, Walter. 1985. "A Solution to Forrester's Paradox of Gentle Murder." *Journal of Philosophy*, 82(3): 162–68. https://doi.org/10.2307/2026353%E2%80%A2

Von Wright, Georg Henrik. 1999. "Deontic Logic: A Personal View." *Ratio Juris*, 12(1): 26–38. https://doi.org/10.1111/1467-9337.00106

Organon F 30 (4) 2023: 323–338
https://doi.org/10.31577/orgf.2023.30402

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

RESEARCH ARTICLE

Discussion Notes on Rettler's Active Reflection

Seong Soo Park*

Received: 12 October 2023 / Revised: 31 October 2023 / Accepted: 2 November 2023

Abstract: According to Rettler, there are three types of control that we should consider in relation to the legitimacy of doxastic blame: Intention-based, reason-based, and influence-based control. Rettler argues that among these, influence-based control is the only type of control that is necessary and sufficient for fulfilling the control condition required for legitimate doxastic blame. The aim of this short discussion paper is to critically assess Rettler's view of doxastic control. By doing so, I attempt to defend the reason-responsiveness view from Rettler's criticism and cast doubt on whether influence-based control has a positive epistemic influence on beliefs.

Keywords: Lindsay Rettler; doxastic blame; ethics of belief; reason-responsiveness.

1. Introduction

Lindsay Rettler (2018) argues that blame imposed on our beliefs can be legitimate by refuting the following argument, what she calls the *anti-blame argument* (Rettler 2018, 2206):

* Sungkyunkwan University
https://orcid.org/0000-0002-7403-6343
(03063) The Department of Philosophy, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongro-gu, Seoul, Republic of Korea
seongsoo.phil@gmail.com

(P1) If agents are legitimately subject to blame for their beliefs, they have control over their beliefs.

(P2) However, agents have no control over their beliefs.

(C) Therefore, agents are not legitimately subject to blame for their beliefs.

One issue of the anti-blame argument is that the meaning of "control" is difficult to pin down. To avoid this difficulty and also to deny (P2) in the argument, Rettler divides control into three types: intention-based, reason-based, and influence-based control. Specifically, Rettler argues that we have influence-based control and that this type of control is both sufficient and necessary for fulfilling the *control condition* for legitimate doxastic blame, which is a requirement for holding oneself or others responsible for their beliefs legitimately.[1] Consequently, she concludes that (P2) is false.[2] Furthermore, she maintains that intention-based control is sufficient but not necessary to satisfy the control condition, while reason-based control is necessary but not sufficient to satisfy it.

This short discussion paper aims to show that Rettler's claims on reason-based control and influence-based control, as presented in her 2018 paper, are not convincing. To show this, the paper is structured as follows: In Sect. 2, I introduce the three types of control proposed by Rettler. Then, in Sect. 3, I evaluate Rettler's claim that reason-based control cannot suffice to fulfill the control condition for legitimate doxastic blame. Finally, in Sect. 4, I investigate whether influence-based control is sufficient for legitimate doxastic blame, and whether it has a positive epistemic effect on beliefs.

[1] According to Rettler (2018, 2207), there are other requirements for legitimate blame, such as the attribution condition, the value condition, or the epistemic condition. However, following Rettler (2018) I will focus in this paper on the issue of which control can fulfill the control condition for legitimate doxastic blame.

[2] Note that some philosophers like Feldman (2000) denies the first premise in the anti-blame argument.

2. Overview of Rettler's three types of control

In this section, I will briefly outline the three types of control suggested by Rettler and introduce her core ideas on them.

To begin with, Rettler (2018, 2211) defines (direct) intention-based control as follows: "An agent has direct intention-based control over ϕ-ing just in case she can ϕ directly as the result of an intention to ϕ." This kind of control is the one that we have over our actions. Surely, I can raise my hands as the result of an intention to do so. However, as Williams (1973) and Alston (1985) point out, we do not have this kind of control over our beliefs.

Rettler agrees that we do not have intention-based control over our beliefs but further argues that this kind of control, even if we had it, cannot be necessary for fulfilling the control condition. To be specific, she points out that there are many cases in which we can legitimately blame ourselves or others for something, even if we or they don't have intention-based control over it. For example, having an unhealthy body or an inconsiderate characteristic is neither a state nor a tendency that can be attained instantly as a direct result of intentional action. Nevertheless, it seems that we can legitimately blame someone for their unhealthy state under the assumption that there is a legitimate reason. Thus, Rettler concludes that having intention-based control is not a necessary condition for legitimate doxastic blame.

The second type of control that Rettler considers is based on the reason-responsiveness view, so it would be better to first briefly introduce the key idea of this view. The key idea lies in the observation of Ryan (2003) and Steup (2008) that both figuring out what to do and what to believe are based on reasons. The difference between our actions and beliefs is merely that we act based on practical reasons, whereas we form and hold beliefs based on epistemic reasons.[3] Thus, according to this view, the reason why we cannot, for example, believe that the Earth is flat by executing our intention to believe it is simply because we do not have a proper reason for holding such a belief: If we had a proper reason, then we would have been

[3] This claim is challenged by pragmatism. See McCormick (2014) and Rinard (2018).

able to believe it. That is, the reason-responsiveness view posits that we have the capacity for forming beliefs in a way that responds properly to reasons. Rettler calls this capacity *reason-based control*.

At first glance, it seems that reason-based control helps explain why some of our practices involving doxastic blame are legitimate. This is because an agent's capacity to properly respond to reasons means that they can hold beliefs that are likely to be true. Thus, if an agent adheres to false beliefs even though they possess the capacity to properly respond to reasons, they may be considered blameworthy or accountable for their false beliefs.

However, Rettler argues that having the capacity for responding to reasons (i.e., having reason-based control) is necessary but not sufficient for fulfilling the control condition required for legitimate doxastic blame. Rettler writes:

> To be able to actively reflect, one must be able to respond to reasons—in the form of both sensitivity to reasons and sensitivity to appreciation of reasons. However, such responsiveness is insufficient on its own to constitute the control required for legitimate doxastic blame. In order to satisfy the control condition, one must additionally be able to execute intentions to carry out the mental actions that constitute reflection. What helps explain the legitimacy of doxastic blame is not our ability to respond to reasons per se, but rather our ability to perform mental actions that influence the process of responding to reasons. These actions make a difference to whether we believe a proposition. Therefore, it would be misleading to characterize our doxastic control as responsiveness to reasons. We ought instead to characterize it in terms of our capacity for active reflection (Rettler, 2018, 2216–17, *italics* added).

As the passage shows, Rettler believes that only the capacity for active reflection helps directly explain the legitimacy of doxastic blame. From her perspective, having the capacity to properly respond to reasons is a mere necessary condition for having the capacity to engage in active reflection, not sufficient for fulfilling the control condition. To argue for this, Rettler (2018: 2215) claims that the capacity for responding to reasons must be further specified into two types of reflection: active and passive reflection.

The distinctive difference between them, according to her, involves whether an agent intends to engage in reflection. Specifically, if reflection stems from a direct result of an intention, it is deemed active. Conversely, if reflection just manifests without any explicit intention, it is deemed passive.

Thus, according to Rettler, active reflection is a mental state that an individual can engage in as a result of an intention to do so, whereas passive reflection is a state that one might find themselves in, often regarded as involuntarily, due to unexpected thoughts and memories that suddenly pop up into their head. Since the notion of passive reflection is important for our discussion but might be unfamiliar, it would be worth looking at how Rettler describes it:

> Consider what it would be for an agent to engage in passive reflection. Anyone who's ever tried to figure out a solution to a complicated problem has likely experienced finding themselves thinking about what they're going to eat for dinner that night…or wondering if they're going to hear back from that job interview two weeks ago…An agent engaged in reflection passively would be in a similar situation with respect to the various mental events that compose reflection…She would find herself wondering whether a certain reason supports believing a proposition, though she didn't intend to direct her attention there. She would suddenly remember a piece of evidence that she'd previously forgotten, and so forth (Rettler 2018, 2215–6).

Based on the above distinction, Rettler (2018, 2216) argues that we can imagine a person who can at best only passively engage in reflection despite having the ability to respond to reasons and, in this case, they cannot be legitimately blamed for their beliefs. (We will see about the reason for this in detail in Sect. 3.) This is why Rettler concludes that reason-based control is not sufficient to satisfy the control condition required for legitimate doxastic blame: reason-based control could involve only passive reflection.

By making the distinction between passive and active reflection explicit, Rettler suggests what she calls the reflective control view, which assumes that we have the capacity to engage in active reflection. The most distinct characteristic of the capacity for active reflection is that it is exhibited as a result of the intention to contemplate. Thus, when an agent performs active

reflection based on their intention, this can indirectly epistemically impact on their beliefs. In other words, the capacity for active reflection is a mental action that can indirectly affect our beliefs. Rettler calls this capacity influence-based control, arguing that it is both necessary and sufficient for satisfying the control condition required for legitimate doxastic blame.

3. Does responsibility always imply the execution of an intention?

As previously mentioned, Rettler claims that reason-based control is insufficient for legitimate doxastic blame. In this section, I will argue that Rettler fails to properly show that this claim is true. I will begin by mentioning the following theses that Rettler explicitly accepts:

(1) Active reflection is the mental action of contemplating something through the execution of an intention, whereas passive reflection is not.

(2) Active reflection is both necessary and sufficient for fulfilling the control condition required for legitimate doxastic blame.

(3) Passive reflection is not sufficient for fulfilling the control condition required for legitimate doxastic blame.

(4) Passive reflection is necessary but not sufficient for having the capacity to reflect actively.

Furthermore, it seems that Rettler also accepts the following, given that she rejects the reason-responsiveness view by contrasting active reflection with passive reflection:

(5) Having the capacity to properly respond to reasons is sufficient for having the capacity for passive reflection.

All five of the theses mentioned can be controversial. Nevertheless, they clearly reflect Rettler's point of view, namely that an agent can be legitimately blamed only for their *intentional actions* (whether they involve physical movement or mental activity) that were carried out with full

awareness and control. This also explains why she does *not* count reason-based control, such as the capacity for passive reflection, as sufficient for fulfilling the control condition required for legitimate doxastic blame. The reason is that passive reflection is not a mental action that the agent can engage in by executing their intention. In other words, from Rettler's perspective, for an agent's actions, or beliefs to be legitimately subject to blame, they must be controllable at least indirectly via the execution of an intention.

For example, a person arriving late to an appointment can be legitimately subject to blame, even though the action of not arriving late to the appointment cannot be achieved based solely on the direct execution of an intention. This is because, as a result of executing an intention, the person could have performed actions, such as waking up earlier, that could have indirectly prevented her from arriving late to the appointment. Similarly, we can be legitimately blamed for holding incorrect beliefs because we have the capacity to engage in a mental action that may indirectly prevent us from holding such beliefs. This is reason why Rettler believes that influence-based control is the only kind of control that is necessary and sufficient for fulfilling the control condition required for legitimate doxastic blame.

However, doxastic compatibilists (e.g., Ryan (2003), Steup (2012), and McHugh (2014)) have refuted ideas similar to Rettler's on a few occasions. Specifically, Ryan (2003) distinguishes intending to φ from doing φ intentionally to argue that an agent who lacks direct control over their beliefs is still responsible for their beliefs. She writes:

> I think consciously deciding to do x or explicitly intending to do x is sufficient, but not necessary, for doing x intentionally. Doing something purposefully is necessary for an action to be done intentionally. And, one can do something unconsciously, automatically, and purposefully (Ryan 2003, 70-71).

That is, according to Ryan, there are things that can be done intentionally without intending to do them.

What motivates Ryan's distinction? Doxastic compatibilists, including Ryan, share the fundamental premise of the reason-responsiveness view: there is no significant difference between action and belief in their nature. (This is why the reason-responsiveness view is regarded as a specific type

of doxastic compatibilism.) Thus, according to this view, most features attributed to action can also be attributed to belief. Ryan emphasizes that sometimes actions can be done intentionally without having an explicit intention to do so and tries to establish the same point for beliefs.

To be specific, consider, for instance, the action of shifting gears from P to D while driving. Such an action is often performed unconsciously without involving any explicit intention. Still, it seems reasonable to classify it as one carried out intentionally. After all, switching gears is not something that is done compulsively, but something that is typically done to achieve a specific goal. The same point can be applied to other actions, such as typing individual letters to compose a passage or brushing one's teeth.

Based on this point, Ryan (2003, 73–74) argues that although believing a specific content is not something we can do through the direct execution of an intention, it is still something we can do intentionally. From her perspective, this is because we have the epistemic purpose of forming (or holding) true beliefs and avoiding false beliefs and try to align our beliefs in line with this purpose. Thus, the epistemic purpose provides a reason to hold true beliefs, which explains why beliefs can typically be regarded as intentional: just as actions carried out for a specific purpose have a reason to be performed and thereby can be considered intentional, beliefs aligned with the epistemic goal can also be considered intentional. Consequently, according to Ryan, one can intentionally hold a belief even if there is no explicit intention behind it.

Note that if Ryan's point is correct, an agent may be held responsible for their beliefs, even if they have not actively reflected on them. This is because beliefs formed solely through passive reflection can still be considered intentional in that such reflection is grounded in an agent's epistemic reasons. Therefore, it could be argued that an agent can be held responsible for their beliefs even in the case where their beliefs were formed solely through passive reflection.

The key issue here does not lie in the truth of Ryan's point but rather lies in the incompleteness of Rettler's attempt to reject the reason-responsiveness view.[4] Rettler argues that an agent who only has the capacity for

[4] In fact, it is controversial whether doxastic compatibilism is tenable. See Nottelmann (2006), Peels (2014), and Booth (2014).

passive reflection and lacks the capacity for active reflection cannot be legitimately blamed for their beliefs. This is because, according to Rettler, passive reflection is not a mental action that can be executed by direct intention. However, Ryan contends that beliefs do not have to be held based on active reflection in order to be considered intentional, and thus can still be regarded as objects of responsibility without being linked to active reflection. Thus, for Rettler to convincingly refute the reason-responsiveness view, she must present an argument that undermines the claim that not only is active reflection intentional, but passive reflection is as well. However, Rettler does not provide such an argument, which leaves her criticism incomplete.

4. Problems with the reflective control view

4.1 Considerations from the Perspective of Empirical Studies

In the previous section, I argued that Rettler's criticism of the reason-responsiveness view is not conclusive. However, this does not guarantee the falsity of the reflective control view. Thus, in this section, I will argue that the reflective control view is not convincing.

The most serious problem with the reflective control view arises from Rettler's claim that influence-based control (i.e., having the capacity for active reflection) is *sufficient* for satisfying the control condition required for legitimate doxastic blame. This is because, in order to justify this claim, Rettler should show that the successful execution of the capacity for active reflection typically has a *positive* epistemic influence on beliefs. Without this, the argument that having influence-based control is sufficient for satisfying the control condition required for legitimate doxastic blame would be unconvincing. The problem is that as Kornblith (2012: 20-26) points out, there are several empirical studies that indicate that active reflection typically has negative effects on an agent's epistemic states (e.g., Garnham and Oakhill, 1994; Stanovich, 1999).

Rettler anticipated that the reflective control view may be criticized in this way. Thus, she suggests two reasons to argue that influence-based control is not the capacity that affects an agent's beliefs in every possible way,

but rather the one that affects an agent's beliefs in a positive way. The two reasons suggested by Rettler (2018, 2221) can be articulated as follows:

> Reason 1: Just as there are several studies that show that active reflection typically has negative effects on an agent's epistemic states, there are also several studies that indicate that it typically has positive effects (e.g., Gagné and Smith (1962); DeWall, Baumeister, and Masicampo (2008); Small, Loewenstein, and Slovic (2007)).
>
> Reason 2: Cases where an agent's active reflection has negative epistemic effects on their beliefs correspond exactly to cases where the agent is free from legitimate criticism of their beliefs.

However, from my perspective, the two reasons mentioned above fail to support Rettler's claim that active reflection typically has a positive epistemic influence on beliefs. First, (Reason 1) can, at best, only show that the empirical studies on active reflection cited by Kornblith are controversial. When empirical studies are contradictory on some issue, we should reserve our judgment on that issue until the supporters of one side win the argument.[5] However, Rettler cannot simply reserve judgement on the effect of active reflection. The success of the reflective control view depends on whether the claim that active reflection has a positive epistemic effect on beliefs is true. Thus, (Reason 1) is not sufficient to support the reflective control view. At least, Rettler needs to show that certain aspects of the studies she cites can help refute the studies Kornblith (2012) mentions.

Second, with regard to (Reason 2), the example of cases in which Rettler thinks doxastic blame cannot be legitimately attributed to an agent is somewhat questionable. The example is as follows:

> For example, suppose Kate incorrectly believes that Colgate is a more effective toothpaste than Crest, but unbeknownst to her she's influenced to believe this by the fact that Colgate toothpastes are located to the right of Crest ones in the shopping aisle.

[5] Someone may argue that even when the empirical evidence is evenly balanced, other factors, such as common sense or inherent understanding, may come into play. I am not motivated to refute this, as it represents a case where one side prevail over the other.

> Assume that if she were to reflect, she would not be able to identify this influence. In this case, it seems unintuitive to think that Kate is legitimately blameworthy for her belief (Rettler 2018, 2220).

Rettler's intuition regarding the above case is that it is not legitimate to attribute doxastic blame to Kate. However, why is such blame not legitimate? Rettler explains that the lack of legitimacy for such blame is due to Kate's inability to recognize her wrong belief that Colgate toothpaste is more efficient than Crest toothpaste, despite engaging in *active reflection*. More specifically, Rettler says the following:

> Do we really think it's legitimate to blame someone for an incorrect psychological belief about their abilities if no amount of reflection could possibly dislodge that belief? I think not. In such a case, it's unfortunate that the person has an incorrect belief, but since she lacks the capacity for causally influential reflection, she cannot acknowledge that the belief is incorrect and thus that she should give it up. Given that she would not be able to respond to the demand inherent in the blame, it would not be legitimate to blame her for the belief (Rettler 2018, 2221, italics added).

Surely, based on the supposition, Kate cannot derive an epistemic benefit from the influence of active reflection in the above case. However, note that this alone does not allow us to conclude that it is not legitimate to blame Kate for her belief. To draw this conclusion, we need the additional premise that the reflective control view is true. This is because, contrary to what Rettler claims, there is an intuition that it is legitimate to attribute doxastic blame to Kate since her belief in question is objectively false.

Regarding this point, Rettler might argue that the mere fact that Kate holds a false belief does not suffice as a legitimate reason to blame her. As highlighted in the cited passage, Rettler believes that an agent must be able to respond to the inherent demands of blame to be legitimately held responsible for their belief. Thus, if we show that Kate has a form of control—distinct from influence-based control—over her beliefs that allows her to meet the demands of blame, it would suffice to refute the proposed objection.

Based on this motivation, I will briefly argue that the reason-responsiveness view can support the idea that Kate's belief can be legitimately blamed. Although Kate is in a state where active reflection cannot have a positive epistemic effect on her belief due to the current location of toothpaste products, as long as she has the capacity for passive reflection (that is, if she can properly respond to epistemic reasons), she would be able to respond properly to doxastic blame if it were suggested.[6] To be more specific, if someone blamed Kate for believing that a particular product is more efficient than another solely based on its location, she would be able to recognize that her belief was based on incorrect reasoning and would be able to try to modify it in response to such blame. Thus, Kate's expected response, contrary to Reltter's claim in the above passage, indicates that Kate has the ability to respond to the inherent demands of blame even if she is in a situation where active reflection cannot have a positive epistemic effect on her belief.

Consequently, cases where active reflection has a negative epistemic effect on beliefs do not match cases where an agent cannot be legitimately blamed for their beliefs. To argue that there is a connection between the former and latter cases, Rettler must presuppose that the reflective control view is true, which would beg the question.

In summary, the reasons suggested by Rettler do not support that the reflective control view is true. The first reason only indicates that the reflective control view is not entirely false. The second reason can be established only if the reflective control view is true and therefore cannot support the claim that the reflective control view is true.

4.2 Considerations from the Conceptual Perspective

As presented above, Rettler claims that active reflection has a positive epistemic influence on beliefs. In this section, I will argue that this claim can also be rejected without relying on empirical studies. To do so, I will

[6] Someone might think that the effect of passive reflection seems dubious. However, it is Rettler who suggests the distinction between passive and active reflections. What I aim to show is that even if her distinction is valid, it does not necessarily follow that Kate is blameless.

begin by clarifying what exactly the phrase "have a positive epistemic influence" means.

One intuitive approach to interpreting this phrase is to connect it with our epistemic goal. It is widely held that our epistemic goal is to form true beliefs and avoid false beliefs. Thus, according to this approach, if Rettler's claim is true, active reflection can be understood as a mental action that enables us to form beliefs that are true and avoid beliefs that are false. However, it is not the case that active reflection can always lead to achieving this epistemic goal. For instance, imagine the inhabitants of a fictional island, say, Epistemia. In Epistemia, people hold deeply-rooted superstitious beliefs, such as the one that thunders occur because of God's anger. With no scientific resources available to disprove these beliefs, the Epistemians would fail to identify their false beliefs, even if they engaged hard in active reflection. This shows that active reflection may not be significantly helpful for an agent in achieving their epistemic goal, especially when the society or community to which they belong operates upon a significant number of false beliefs.

The lesson to be drawn from the above case is that, for an agent to experience a positive epistemic effect on their beliefs through active reflection, they must already have a significant number of true beliefs in their current belief system. I will call this necessary condition for having a positive epistemic effect from active reflection the *robust belief system condition*. Note that performing active reflection is irrelevant to satisfying the robust belief system condition. Rather satisfying a robust belief system condition relies on, for example, having a stable higher-order recognition state, such as accurately seeing and hearing, being sober, or being capable of various types of epistemic behaviors that allow an agent to gather additional evidence. Thus, if the phrase "have a positive epistemic influence" is interpreted in the manner described above, Rettler's claim will be false.

Rettler might object that I have interpreted the phrase "have a positive epistemic influence" too strongly. To be specific, she may appeal to Burge's explanation of (active) reflection to clarify the meaning of the phrase. According to Burge (1996), reflection is a mental action that can affect second-order beliefs when first-order beliefs are modified by changes in environment (or by additional information). Thus, Rettler could argue that active

reflection has a positive epistemic effect in the sense that it can rectify our second-order beliefs when our existing first-order beliefs turn out to be incorrect.

However, if active reflection has a positive epistemic effect on beliefs just in the sense described above, Rettler's claim that influence-based control is sufficient for fulfilling the control condition required for legitimate doxastic blame faces difficulties. To see this, consider the following case: Betty was trying to solve a difficult math problem, but she wasn't sure about a crucial principle needed to solve it. Instead of looking up the principle in a book (by using a library), she relied only on her reflection, even though she knew she might not remember it correctly. She submitted her answer and ultimately failed to solve the problem correctly.

In this case, it seems clear that we can hold Betty responsible for having failed to solve the math problem. However, note that we cannot blame Betty simply for not actively reflecting. In other words, the mere act of not actively reflecting about something is not enough to attribute blame to Betty. She did reflect on the content of the principle, and her reflection had a positive epistemic effect *in the sense* that if she had examined a book on mathematics in a library, she could have rectified her false beliefs on the math problem. Thus, the key to doxastic blaming Betty is that she had the ability to gather additional evidence, which could have satisfied the robust belief system condition, but she did not exert this ability. Consequently, if active reflection is interpreted in this rather weak sense, it will not be sufficient for fulfilling the control condition required for legitimate doxastic blame.

In summary, regardless of how we interpret active reflection, whether in terms of its contribution to our epistemic goal or its ability to derive true second-order beliefs, Rettler's claim about its positive influence faces challenges.

5. Conclusion

Rettler argues that reason-based control is insufficient for satisfying the control condition required for legitimate doxastic blame, and that only influence-based control is both a sufficient and necessary condition for

fulfilling it. However, her claim about reason-based control is not conclusive since she has not considered whether passive reflection is something we can do intentionally. Additionally, the reflective control view is not promising because the central claim that active reflection typically has a positive epistemic effect on beliefs is not well supported. Rettler not only has difficulties providing good reasons that can refute the empirical studies that threaten the central claim but also has difficulties embracing it under the two possible interpretations.

Acknowledgements

I am deeply thankful to an anonymous reviewer for their valuable comments and suggestions. Special appreciation goes to Inchul Yum whose detailed comments and suggestions resulted in many significant improvements in this paper.

References

Alston, William P. 1985. "Concepts Of Epistemic Justification." In *Epistemic Justification: Essays in the Theory Of Knowledge*, edited by W. Alston, 81–114. Ithaca: Cornell University Press

Booth, Anthony Robert. 2014. "On Some Recent Moves in Defence of Doxastic Compatibilism." *Synthese* 191 (8): 1867 –80. https://doi.org/10.1007/s11229-013-0378-x

Burge, Tyler. 1996. "Our Entitlement to Self-Knowledge." *Proceedings of the Aristotelian Society* 96 (1): 91 –116. https://doi.org/10.1093/aristotelian/96.1.91

DeWall, Nathan. C., Baumeister, Roy, & Masicampo, E. J. 2008. "Evidence that Logical Reasoning Depends on Conscious Processing." *Consciousness and Cognition* 17 (3): 628–45.

Feldman, R. 2000. "The Ethics of Beliefs." *Philosophy and Phenomenological Research* 60 (3): 667–95. https://doi.org/10.2307/2653823%E2%80%A2

Gagné, Robert. M., & Smith, Ernest C. 1962. "A Study of the Effects of Verbalization on Problem Solving." *Journal of Experimental Psychology* 63 (1): 12–18. https://doi.org/10.1037/h0048703

Garnham, Alan, & Oakhill, Jane. (1999). *Thinking and Reasoning*. Oxford: Blackwell.

Kornblith, Hilary. 2012. *On Reflection*. Oxford: Oxford University Press.

McCormick, Miriam. 2014. *Believing against the Evidence: Agency and the Ethics of Belief*. New York: Routledge.

McHugh, Conor. 2014. "Exercising Doxastic Freedom." *Philosophy and Phenomenological Research* 88 (1): 1–37. https://doi.org/10.1111/j.1933-1592.2011.00531.x

Nottelmann, Nikolaj. 2013. "The Analogy Argument for Doxastic Voluntarism." *Philosophical Studies* 131 (3): 559–82. https://doi.org/10.1007/s11098-004-7489-7

Peels, Rik. 2014. "Against Doxastic Compatibilism." *Philosophy and Phenomenological Research* 89 (1): 679–702. https://doi.org/10.1111/phpr.12040

Rettler, Lindsay. 2018. "In Defense of Doxastic Blame." *Synthese* 195 (5): 2205-26. https://doi.org/10.1007/s11229-017-1332-0

Rinard, Susanna. 2018. "Believing for Practical Reasons." *Noûs* (4): 763–84. https://doi.org/10.1111/nous.12253

Ryan, Sharon. 2003. "Doxastic Compatibilism and the Ethics of Belief." *Philosophical Studies* 114 (1-2): 47–79. https://doi.org/10.1023/A:1024409201289

Small, Deborah A., Loewenstein, George., & Slovic, Paul. 2007. "Sympathy and Callousness: The Impact of Deliberative Thought on Donations to Identifiable and Statistical Victims." *Organizational Behavior and Human Decision Processes* 102 (2): 143–53. https://doi.org/10.1016/j.obhdp.2006.01.005

Stanovich, Keith. 1999. *Who Is Rational? Studies of Individual Differences in Reasoning.* Lawrence Erlbaum.

Steup, Matthias. 2008. "Doxastic Freedom." *Synthese* 161 (3): 375–92. https://doi.org/10.1007/s11229-006-9090-4

Steup, Matthias. 2012. "Belief Control and Intentionality." Synthese 188 (2): 145–63. https://doi.org/10.1007/s11229-011-9919-3

Williams, Bernard. 1973. "Deciding to Believe." In *Problems of the Self*, edited by B. Williams Cambridge: Cambridge University Press: 136–51. https://doi.org/10.1017/CBO9780511621253.011

Organon F 30 (4) 2023: 339–349
https://doi.org/10.31577/orgf.2023.30403

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

RESEARCH ARTICLE

Removing an Inconsistency from Jago's Theory of Truth

Nathan William Davies*

Received: 31 July 2023 / Revised: 30 November 2023 / Accepted: 4 December 2023

Abstract: I identify an inconsistency in Jago's theory of truth. I show that Jago is committed to the identity of the proposition that the proposition that A is true and the proposition that A. I show that Jago is committed to the proposition that A being true because A if the proposition that A is true. I show that these two commitments, given the rest of Jago's theory, entail a contradiction. I show that while the latter commitment follows from Jago's theory of truth, even by Jago's own lights he should not be committed to the identity of the proposition that the proposition that A is true and the proposition that A.

Keywords: Truth; truthmaking; proposition; state of affairs; explanation; redundancy.

1. Preliminaries

Jago thinks that to be true is to be made true by something (2018, e.g. 1, 74, 81). He thinks that the primary bearers of truth and falsity are propositions (2018, e.g. 5, 235, 316). He thinks that propositions are sets of their truthmakers ((2017); (2018, e.g. 249–252); (2022)). Ignoring subtle

* Kenilworth
https://orcid.org/0000-0003-4636-9201
Kenilworth, United Kingdom
nwd.correspondence@gmail.com

Jago exegesis,[1] he thinks that a truthmaker for a proposition is a state of affairs which, were it to exist, would make the proposition true ((2018, 81, 261–266); (2022, 231)). I take these to be Jago's core commitments regarding truth and propositions.

A note on notation: 'A' is used, as in (Jago 2018), as a variable or meta-variable; ⌜⟨A⟩⌝ is an abbreviation of ⌜the proposition that A⌝; ⌜|A|⌝ is an abbreviation of ⌜the state of affairs that A⌝; 'P' is an abbreviation of 'Prague is the capital of the Czech Republic'; and '[' and ']' are used for editing and supplementing text.

2. Contradiction!

On the basis of what Jago has written, he is committed to the following argument being sound:[2]

1. ⟨⟨P⟩ is true⟩ = ⟨P⟩. [See §2.1 and §3]
2. ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true. [⟨⟨P⟩ is true⟩ is true & Jago's theory]
3. ⟨P⟩ is true because ⟨P⟩ is true. [1, 2, & Jago's theory]
4. ⟨P⟩ is true but not because ⟨P⟩ is true. [Assumption]

⊥ [3 & 4]

The argument cannot be sound, of course, because its conclusion is a contradiction. In the following four subsections I explain why Jago is committed to the truth of each of the premises. In §3 I show that even by Jago's own lights he should never have committed himself to the truth of Premise 1.

[1] For those interested in such things, consult e.g. (Jago 2017, 295) and (Jago 2018, 250, 261–266).

[2] Similar arguments, not directed at Jago specifically, have been given by Trueman (2022, 104) and Künne (2018, 51).

2.1 Premise 1

Jago commits himself to the truth of Premise 1 in the following passage from *What Truth Is*:

> Given our analysis of *being true*, ⟨⟨A⟩ is true⟩ is identical to the existential proposition ⟨∃*x x* truthmakes ⟨A⟩⟩, which we identify with the set of all its witnesses: all truthmakers for ⟨A⟩. So ⟨⟨A⟩ is true⟩ is identical to ⟨A⟩ itself. (Jago 2018, 251–52)

2.2 Premise 2

Regarding Premise 2, consider the following passage from *What Truth Is*:

> Here's [a] platitude about truth, governing our use of 'is true': [if ⟨A⟩ is true, then] ⟨A⟩ is true because A (Wright 1992, 25–6; Lynch 2001, 747). (Or rather, any instance of that scheme is a platitude, wherever ⟨A⟩ is a proposition.) … underlying the monadic existential property *having a truthmaker* is the binary relation — *truthmakes* —. An entity *x* stands in its first argument place when the proposition in the second argument place is true in virtue of *x*'s existence. This 'in virtue of' is used canonically to express metaphysical explanations. We may replace 'is true in virtue of *x*'s existence' with 'is true because *x* exists'. And there we have our explanation for the platitude… (Jago 2018, 75–76)

Jago is committed to the following proposition being true:

If ⟨⟨P⟩ is true⟩ is true, then ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true.

Given that Prague is the capital of the Czech Republic, ⟨P⟩ is true. Given that ⟨P⟩ is true, ⟨⟨P⟩ is true⟩ is true. Hence Jago is committed to the truth of Premise 2:

2. ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true.

One might wonder whether Jago should have been committed to the truth of Premise 2. I think, partly on the basis of the passage just quoted, that Jago's theory of truth and ⟨⟨P⟩ is true⟩ entail the truth of Premise 2; in order to explain why I think this, I shall present an argument the soundness

of which I think Jago was committed to and whose conclusion, given the truth of ⟨⟨P⟩ is true⟩, entails Premise 2. In formulating this argument, I have benefitted from Liggins' understanding of the quoted passage (Liggins 2019).

There are five suppositions under which the conclusion shall be derived.

Supposition I

If ⟨⟨P⟩ is true⟩ is true, then something makes ⟨⟨P⟩ is true⟩ true.

Jago thinks that if something is true then it is made true by something (2018, ch.3). So Jago is committed to the truth of this supposition. In fact, Jago thinks that for something to be true just is for it to be made true by something (§1). So he is definitely committed to the truth of this supposition.

Supposition II

If something makes ⟨⟨P⟩ is true⟩ true, then |⟨P⟩ is true| makes ⟨⟨P⟩ is true⟩ true.

That Jago is committed to this supposition, and analogous suppositions, is implicit in (Liggins 2019). Given that Jago believes in states of affairs (2018, §§2.5–2.6), how could he not be committed to |⟨P⟩ is true| making ⟨⟨P⟩ is true⟩ true?

Supposition III

If |⟨P⟩ is true| makes ⟨⟨P⟩ is true⟩ true, then ⟨⟨P⟩ is true⟩ is true because |⟨P⟩ is true| exists.

Jago commits himself to this supposition in the passage quoted above. Additionally, Jago's theory entails that if a state of affairs makes a proposition true, that proposition is true in part because that state of affairs exists and in part because that state of affairs is a member of it (2018, 75–76, 250, 253).

Supposition IV

If |⟨P⟩ is true| exists, then |⟨P⟩ is true| exists because ⟨P⟩ is true.

Again, this supposition is implicit in (Liggins 2019). As is:

Supposition V

If ⟨⟨P⟩ is true⟩ is true because |⟨P⟩ is true| exists and |⟨P⟩ is true| exists because ⟨P⟩ is true, then ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true.

It seems to be necessary for Jago to accept both of these suppositions, if he is to explain why if ⟨⟨P⟩ is true⟩ is true, then ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true.

Here is the argument:

i. ⟨⟨P⟩ is true⟩ is true. [Assumption]

ii. Something makes ⟨⟨P⟩ is true⟩ true. [i & Supposition I]

iii. |⟨P⟩ is true| makes ⟨⟨P⟩ is true⟩ true. [ii & Supposition II]

iv. ⟨⟨P⟩ is true⟩ is true because |⟨P⟩ is true| exists. [iii & Supposition III]

v. |⟨P⟩ is true| exists. [iv]

vi. |⟨P⟩ is true| exists because ⟨P⟩ is true. [v & Supposition IV]

vii. ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true. [iv, vi, & Supposition V]

If ⟨⟨P⟩ is true⟩ is true, then ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true. [i–vii]

Hence Jago's theory and ⟨⟨P⟩ is true⟩ entail Premise 2:

2. ⟨⟨P⟩ is true⟩ is true because ⟨P⟩ is true.

2.3 Premise 3

If Jago is committed to the truth of Premise 1 and he is committed to the truth of Premise 2, then he is committed to the truth of Premise 3. I will show this by showing that Jago is committed to the soundness of the following argument:

a. Premise 2 differs in truth-value from Premise 3 only if the explanandum of Premise 2 differs from the explanandum of Premise 3. [Assumption]

b. The explanandum of Premise 2 differs from the explanandum of Premise 3 only if |⟨⟨P⟩ is true⟩ is true| differs from |⟨P⟩ is true| or ⟨⟨⟨P⟩ is true⟩ is true⟩ differs from ⟨⟨P⟩ is true⟩. [Assumption]

c. ⟨⟨P⟩ is true⟩ = ⟨P⟩. [Premise 1]

d. |⟨⟨P⟩ is true⟩ is true| = |⟨P⟩ is true|. [c & Jago's theory]

e. ⟨⟨⟨P⟩ is true⟩ is true⟩ = ⟨⟨P⟩ is true⟩. [c & Jago's theory]

f. The explanandum of Premise 2 is identical to the explanandum of Premise 3. [b, d, & e]

Premise 2 does not differ in truth-value from Premise 3. [a & f]

(I benefitted from (Künne 2018, 51–55) in trying to spell out this argument.)

As far as I can tell, every account of states of affairs Jago considers (2018, ch.4) is such that were he to accept it, he would be committed to Premise d following from Premise c (see also (Jago 2022, 238–39)).

The question to ask then is whether Premise e really does follow from Premise c and Jago's theory. The answer is it does. Jago thinks that what is said/stated by someone on a given occasion is a proposition (2018, 236, 238); with that said, consider:

> In saying that David Jones changed the world, I thereby say that David Bowie did, 'Bowie' being the name Jones adopted. (Jago 2017, 294)

> [W]e can contrast what is said with the particular *way* in which it is said. To bring out the idea, suppose Anna and Bob are arguing, Anna insisting that the planet now visible is Hesperus, whereas Bob insists that it is Phosphorus. There is clearly a sense in which they are not really disagreeing at all, for they are both correctly identifying the planet they see. Someone in the know may interject, 'stop arguing, you are saying the same thing!' Nevertheless, both parties are genuinely informed when they come to learn that the planet is correctly called both 'Hesperus' and 'Phosphorus'. What they lacked was *a posteriori* knowledge, not linguistic competence. This shows that the notion of *what is said* in an utterance does not align with the meaning of the

> utterance, or with the speaker's beliefs, or with common knowledge in the conversation. (Jago 2017, 303)

See also (Jago 2018, 236–237, 266–267).

Jago does deny that propositions are 'objects' of belief, but it is clear what he thinks about their identity conditions:

> One application for which the truthmaker approach [to propositions] is (probably) *not* suitable is the analysis of attitude reports. A truthmaker, as commonly understood, is a worldly entity such as a state of affairs. Truthmakers do not, in general, involve modes of presentation. What makes it true that George Eliot wrote *Middlemarch* also makes it true that Mary Anne Evans wrote *Middlemarch*, for they were one and the same person. So (assuming the necessity of identity), the proposition that George Eliot wrote *Middlemarch* will be identical to the proposition that Mary Anne Evans wrote *Middlemarch*, according to the truthmaker approach. Yet one can believe that George Eliot wrote *Middlemarch* without believing that Mary Anne Evans wrote *Middlemarch*. It seems that attitude reports are beyond the scope of the truthmaker approach. (Jago 2022, 238–39)

See also (Jago 2018, 236–238). So it is clear that Jago is committed to Premise e following from Premise c.[3]

Couldn't Jago argue that Premise b is false? Couldn't Jago argue that explananda, like beliefs, are to be individuated more finely than both states of affairs and propositions? He could, perhaps, but in that case, would the putatively metaphysical fact that ⟨P⟩ is true because P really be a fact about ⟨P⟩ as opposed to our conception of it? I assume here that Premise b is true.

[3] Of course, given that Jago committed himself to ⟨⟨A⟩ is true⟩ being identical to ⟨A⟩ (§2.1), he committed himself to ⟨⟨⟨P⟩ is true⟩ is true⟩ being identical to ⟨⟨P⟩ is true⟩. But I wanted to show that even if Jago hadn't committed himself to ⟨⟨A⟩ is true⟩ being identical to ⟨A⟩, and had merely committed himself to ⟨⟨P⟩ is true⟩ being identical to ⟨P⟩, his theory of propositions would have ensured his commitment to the identity of ⟨⟨⟨P⟩ is true⟩ is true⟩ and ⟨⟨P⟩ is true⟩.

Jago is committed to Premise 2 not differing in truth-value from Premise 3. Hence Jago is committed to Premise 3 being true if he is committed to Premise 2 being true. As we have already seen, he is committed to Premise 2 being true. So, Jago is committed to the truth of Premise 3.

2.4 Premise 4

While there is at least one truth whose truth is explained, at least in part, by its being true, namely ⟨at least one proposition is true⟩,[4] there is surely at least one truth whose truth is not explained, not even in part, by its being true. I assume that ⟨P⟩ is such a truth. Jago can deny this of course, but there must be an analogous argument the fourth premise of which Jago couldn't reject. So let us pretend that Jago is committed to the truth of Premise 4.

3. Rejecting Premise 1

Jago is committed to the truth of Premise 3 (because he is committed to the truth of Premise 1 and Premise 2) and he is committed to the truth of Premise 4. Premise 3 and Premise 4 are contradictory, so which premises should Jago not be committed to? Jago shouldn't be committed to the truth of Premise 3 and should never have committed himself to the truth of Premise 1. Let us look again at the quote from §2.1:

> Given our analysis of *being true*, ⟨⟨A⟩ is true⟩ is identical to the existential proposition ⟨∃*x* *x* truthmakes ⟨A⟩⟩, which we identify with the set of all its witnesses: all truthmakers for ⟨A⟩. So ⟨⟨A⟩ is true⟩ is identical to ⟨A⟩ itself. (Jago 2018, 251–252)

Jago made a mistake. To make it clear what the mistake is, let us look at an argument Jago could have offered for Premise 1 which is based on what

[4] ⟨at least one proposition is true⟩ is true because at least one proposition is true. At least one proposition is true at least in part because ⟨at least one proposition is true⟩ is true. I think it follows that ⟨at least one proposition is true⟩ is true at least in part because ⟨at least one proposition is true⟩ is true. See (Krämer 2020) for further discussion.

he says in that passage. Assuming that |P| is the sole truthmaker for ⟨P⟩,[5] the argument is as follows:

α. ⟨⟨P⟩ is true⟩ is identical to ⟨something makes ⟨P⟩ true⟩.

Jago is committed to Premise α because he wrote: "⟨⟨A⟩ is true⟩ is identical to the existential proposition ⟨∃*x* *x* truthmakes ⟨A⟩⟩" (2018, 251).

β. ⟨something makes ⟨P⟩ true⟩ is identical to {|P|}.

He is committed to Premise β because he wrote: "we identify [the existential proposition ⟨∃*x* *x* truthmakes ⟨A⟩⟩] with the set of all its witnesses: all truthmakers for ⟨A⟩" (2018, 251).

γ. ⟨P⟩ is identical to {|P|}.

He is committed to Premise γ because he wrote: "propositions are sets of entities…and we think of those entities as truthmakers for that proposition" (2018, 250). See also (Jago 2022, 231).

δ. ⟨something makes ⟨P⟩ true⟩ is identical to ⟨P⟩.

Premise δ follows from Premise β & Premise γ.

⟨⟨P⟩ is true⟩ is identical to ⟨P⟩.

The conclusion, Premise 1, follows from Premise α & Premise δ.

The argument is valid. Of the three rejectable premises—Premise α, Premise β, and Premise γ—Jago can only reject Premise β: Premise α and Premise γ are entailed by his core commitments concerning truth and propositions (§1).

Jago should not have been committed to Premise β. Jago writes as an introduction to his truthmaker account of propositions:

> The *truthmaker account of propositions* identifies a proposition with the set of its possible truthmakers. That is, a proposition ⟨A⟩ is a set of possible entities, each of which makes it true that

[5] This assumption is consonant with what Jago has said in print (2022, 231), but a different assumption regarding the truthmaker(s) for ⟨P⟩ could be made, if necessary, in the course of giving an analogous argument which would be harder to assess.

> A (or would do, were it to exist). The proposition *that I am sitting* is the singleton containing the state of affairs *that I am sitting*. The proposition *that someone is sitting* is the set containing all possible states of affairs of the form *that x is sitting*. (Jago 2022, 231)

Based on what Jago says there, he is committed to ⟨something makes ⟨P⟩ true⟩ having ||P| makes ⟨P⟩ true| as a member. Assuming that ||P| makes ⟨P⟩ true| is not identical to |P|, ⟨something makes ⟨P⟩ true⟩ is not identical to {|P|}. Hence Premise β is false. Jago should welcome this: if he rejects Premise β, as it seems he must do, he is free to reject Premise 1, as he must do.

Jago's mistake was committing himself to the identity of ⟨something makes ⟨P⟩ true⟩ with the set of its *witness* states. Even if every witness state of ⟨something makes ⟨P⟩ true⟩, i.e. every truthmaker for ⟨P⟩, is a truthmaker for ⟨something makes ⟨P⟩ true⟩, not every truthmaker for ⟨something makes ⟨P⟩ true⟩ is a witness state of it, i.e. not every truthmaker for ⟨something makes ⟨P⟩ true⟩ is a truthmaker for ⟨P⟩. There is an *instance* state of ⟨something makes ⟨P⟩ true⟩, namely ||P| makes ⟨P⟩ true|, which is not a witness state of it, which makes ⟨something makes ⟨P⟩ true⟩ true, but which does not make ⟨P⟩ true.

Jago seems to have made a similar mistake (as I see it) in his discussion of alethic states of affairs. He wrote:

> In general, existential states of affairs are grounded by their instances, and since |⟨A⟩ is true| is identified with |∃*y*(*y truthmakes* ⟨A⟩)|, it follows that its possible grounds are all and only the possible truthmakers for ⟨A⟩.[6] (Jago 2018, 195)

But ||P| makes ⟨P⟩ true| is not a truthmaker for ⟨P⟩, and yet it is a ground for |something makes ⟨P⟩ true|.[7]

[6] Abbreviations have been made consistent with those in this paper.

[7] Incidentally, this shows that Jago's truth-grounding principle, the principle he argues for in §6.3 of (Jago 2018), is false.

4. Conclusion

In conclusion, I have identified an inconsistency in Jago's theory of truth (§2) and I have shown how to remove it without relinquishing Jago's core commitments (§3). The lesson is that Jago cannot serve two masters: he must choose Bolzano over Frege (Künne 2018).

Acknowledgements

I thank an anonymous reviewer for taking the time to read the paper. I thank the editors for giving me a chance to improve it. I thank Prof. Mark Jago for his advice on how to present my main argument, which helped me make the paper much more readable. I thank Maria Zanella for proofreading my drafts.

References

Jago, Mark. 2017. "Propositions as Truthmaker Conditions." *Argumenta* 2 (2): 293–308. https://doi.org/10.23811/47.arg2017.jag

Jago, Mark. 2018. *What Truth Is*. Oxford University Press. https://doi.org/10.1093/oso/9780198823810.001.0001

Jago, Mark. 2022. "Truthmaker Accounts of Propositions." In *The Routledge Handbook of Propositions*, edited by Adam Russell Murray, and Chris Tillman, 231–243. Routledge. https://doi.org/10.4324/9781315270500-16

Krämer, Stephan. 2020. "Puzzles." In *The Routledge Handbook of Metaphysical Grounding*, edited by Michael J. Raven, 271–282. Routledge. https://doi.org/10.4324/9781351258845-25

Künne, Wolfgang. 2018. "Truth, Ascriptions of Truth, and Grounds of Truth Ascriptions." In *Eva Picardi on Language, Analysis and History*, edited by Annalisa Coliva, Paolo Leonardi, and Sebastiano Moruzzi, 31–66. Palgrave Macmillan. https://doi.org/10.1007/978-3-319-95777-7_3

Liggins, David. 2019. "Mark Jago, *What Truth Is*, Oxford University Press, 2018, 356pp., $65.00 (hbk), ISBN 9780198823810." *Notre Dame Philosophical Reviews*. https://ndpr.nd.edu/reviews/what-truth-is/

Trueman, Robert. 2022. "Truth without Dependence." *Aristotelian Society Supplementary Volume* 96 (1): 89–121. https://doi.org/10.1093/arisup/akac012

Organon F 30 (4) 2023: 350–371
https://doi.org/10.31577/orgf.2023.30404

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

RESEARCH ARTICLE

Empirical Inconsistencies Defying Simulationism

Saskia Janina Neumann*

Received: 26 September 2023 / Revised: 8 December 2023 /
Accepted: 14 December 2023

Abstract: In our common understanding, remembering and imagining are two different entities. Yet, with brain research progressing, this common understanding of remembering and imagining changes significantly. Simulationists go as far as to claim that remembering and imagining only differ in their temporal orientation but are part of the same system. In what follows, I want to defend our common understanding of how to distinguish between remembering and imagining. With the help of empirical studies, I will defend the view that remembering and imagining are significantly different and not only different in their temporal orientation. I will base my argumentation on empirical studies which are suggestive of simulationism having gotten it wrong. In this paper, I will firstly introduce the two opposing views of simulationism and the causal theory of memory. With the help of empirical studies, I will secondly show that simulationism faces significant evidence of being wrong and thirdly, will suggest that a slightly changed version of the causal theory of memory does a better job in explaining the introduced research results.

Keywords: Causal Theory of Memory; epistemology of memory; memory traces; philosophy of Memory; philosophy of cognitive science; simulationism.

* Eötvös Loránd University, Budapest
https://orcid.org/0009-0003-8777-4534
Institute of Philosophy, Eötvös Loránd University, Múzeum krt. 4-6, 1088 Budapest, Hungary
SaskiaJNeumann@hotmail.com

1. Introduction

> When I was a child in Canada, I drove with my family through part of the Northwest Territories. At some point, we stopped near a buffalo roll – an area where buffalo roll in the dirt – and my mother told me to stand in it while she took a picture. Naturally, I was frightened, imagining that a buffalo might appear at any moment, set about rolling, and thereby crush me. Or so I seem to remember. In reality, given my age at the time, and given that my parents repeated the amusing story to me a number of times afterwards, I can't be sure that much – or even any – of the content of my apparent memory of the episode actually originates in my experience, as opposed to the subsequent accounts provided by my parents and my own subsequent imaginings of the episode (Michaelian 2016a, 238).

We sure all have memories like these in which we are not completely sure whether we want to call them memories or whether we would rather claim that we only imagine that we have remembered something as somebody else has told us about our supposed memory so often. Encounters like these, however, beg the question of whether there is a clear difference between remembering and imagining to be had. In our common understanding, we would probably defend the view that there is a clear distinction between the two; we must have experienced what we remember. This view very roughly summarizes the causal theory of memory by Martin and Deutscher (1966). Yet, what needs to be added for Martin and Deutscher's causal theory to be causal is a so-called memory trace. A memory trace provides us with a causal link between an experienced event and the representation of the same event at a later time and enables us to remember in the first place as it is the source for what we retrieve once we try to remember something which has happened in our past (Martin – Deutscher 1966). Yet, by now, there is an opposing view to the causal theory of memory; it is simulationism, introduced by Michaelian (2016a). Michaelian does not see the necessity for a causal link between experienced and remembered event. As long as we are able to represent an experienced event, it does not matter where this representation comes from. It might not come from memory directly but might be simulated as well and still be a memory. This

is why Michaelian suggests that imagining the past is remembering and that remembering and imagining thus do not differ in kind. If imagining the past is remembering, we do not necessarily have to have experienced the event we now imagine.

In this paper, I want to argue against simulationism on empirical grounds. In what follows, I will show that simulationism contrasts with empirical studies. I will firstly introduce Martin and Deutscher's theory of memory traces and Michaelian's theory of simulationism. Secondly, I will show studies which suggest that there is a difference between remembering and imagining and thirdly, will argue for these studies showing that this difference is big enough to distinguish between remembering and imagining in kind and not just in degree. Fourthly, I will suggest what alternatives we have if simulationism should be wrong. I will argue that we should consider a version of the causal theory of memory to explain the difference between remembering and imagining. Yet, I suggest that the distinction between remembering and imagining should not be made by the mere presence of a memory trace but by whether a memory trace is solely operative in representing an experienced event.

2. Causal theory of memory and simulationism

As it will be central to my argument in this paper, let me introduce the term 'episodic memory' first. Episodic memory refers, roughly, to the form of memory responsible for allowing us to revisit specific episodes or events from the personal past. It is contrasted with semantic memory, which allows us to recall facts without necessarily giving us access to the episodes in which they were learned' (Michaelian 2016a, 5). Moreover, episodic memory, in contrast to semantic memory, does not only give us propositional contents such as 'A dinosaur is a fossil reptile of the Mesozoic era', which can simply be evaluated as true or false, but, because episodic memory gives us richer representations, it can provide us with more or less accurate representations of what we have experienced. For example, when I remember my encounter with the statue of an avimimus at a dinosaur park, I might remember what its statue looked like but might have forgotten where exactly it was located (was it next to the T-rex or closer to the

Brachiosaurus?), I might remember what it felt like to see this huge creature and also how it felt like when I touched it. This is memories of events and us experiencing them and they thus belong to episodic memory. Yet, if we were to put them into a more factual, propositional form such as: 'Touching the statue of an avimimus feels funny is it is so cold' or 'The avimimus statue was located next to the T-rex', we have semantic memories.

In the case of the buffalo roll, our narrator has a semantic memory if they just recall the proposition 'I was afraid a buffalo would come by and crush me' and an episodic memory if they represent what has happened to them in the specific situation of standing in the buffalo roll such as their feeling of being frightened, how the buffalo roll looked like, what clothes their parents wore etc.. While an episodic memory can thus come with different degrees of accuracy, what both forms of memory have in common, at least in our common understanding of the word remembering, is that the memory of our narrator in the buffalo roll must originate in their own experience. If they just represent having been afraid standing in the buffalo roll based on the testimony of their parents, they do not remember but imagine having been afraid, since what they represent is not caused by what they have experienced but by their parents' testimony.

This common understanding of remembering is captured by Martin and Deutscher's (1966) causal theory of memory. In their understanding of how memory works, we remember once we have a so-called memory trace present. A memory trace is a state or set of states which was produced by a past experience and forms a structural analogue of the experienced event within our memory. It works similar to the grooves in a gramophone record. While the number of wiggles per unit length in a groove determine the pitch of the played music, an experienced event would analogously leave traces in our memory which make it eventually possible to recall an experienced event and represent what has happened at a later time (Martin & Deutscher 1966, 191). The causal connection between experienced and recalled event a memory trace provides is, however, neither a necessary nor a sufficient but an operative condition[1]. A condition is operative if it produces another

[1] A memory trace may still be a necessary and even a sufficient condition for remembering but what is of importance in their paper is the operative condition (Martin & Deutscher 1966, 179).

condition but, in contrast to necessary cases, the caused condition could also have been caused by another condition (Martin & Deutscher 1966, 179). For instance, if I spill a drink, I am the one who has caused the table to be wet. However, my spilling a drink is not necessary for the table to be wet. The drink might as well have been spilled by someone else or the table could have gotten wet for other reasons. While me spilling the drink is not a necessary condition for the table being wet, me spilling the drink is an operative condition as, in this situation, it was me who has spilled the drink and caused the table to be wet. In the case of memory, a memory trace is only an operative and not a necessary condition because the representation of what we have experienced is not necessarily caused by a memory trace. In principle, there could also be other causes for why we represent something, for instance, having spilled a drink. If we represent having spilled a drink, we could be caused by our friend to represent that we have spilled a drink. In this case, we have experienced having spilled a drink and are able to represent that we did so. The operative condition in this case is the testimony of our friend. However, Martin and Deutscher (1966) explain that we do not remember having spilled a drink if the testimony of our friend is the operative condition. In order to remember, it is the experience of having spilled the drink which causes us to represent that we spilled the drink. The testimony of our friend won't do.[2]

The connection between the experience of having spilled a drink and representing us having spilled a drink is a memory trace. A memory trace captures the demand of our common understanding of what remembering is. We need to have experienced an event ourselves and, once we recall the experienced event, this representation needs to be caused by us having experienced that event. A memory trace ensures that this is the case. If there is a memory trace present, we must have experienced an event ourselves and must be able to represent what has happened based on that memory trace, based on the causal connection between experienced event and the representation of that experienced event at a later time. In case of our

[2] As I wanted to explain the operative condition in this paragraph, I have left out the accuracy condition. Yet, keep in mind, that the accuracy condition also needs to be present if we want to be said to remember according to Martin and Deutscher (1966).

narrator's parents' testimony, there is no memory trace present if they did not experience the event of the buffalo roll themselves but only heard about what they have experienced from their parents. Without a memory trace, however, there is no remembering. In case of our narrator's parents' testimony, the narrator can imagine having been afraid of the buffalos crushing them but cannot remember.[3,4,5]

However, even though the causal theory of memory seems to capture our common understanding of remembering, the causal theory of memory, according to Michaelian (2016a, 98f.), does not stand up to recent results

[3] One could object here and ask about cases in which our friend's testimony triggers the activation of a memory trace. As long as the memory trace itself is the operative condition for us remembering, though, we can be said to remember. If our friend's testimony alone triggered the representation of us having spilled a drink, however, the testimony of our friend would be the operative condition and, according to Martin and Deutscher (1966, 179), we could not be said to remember.

[4] Another objection to Martin and Deutscher's operative condition could be made by asking the question of what if another memory trace than the one formed during the experienced event is reponsible for us being able to retrieve an experienced event. If we follow Martin and Deutscher here, the answer is clear: no causal connection between remembered and experienced event means no remembering.

[5] Martin and Deutscher also introduce the term of remembering-how. Remembering-how is not remembering a specific event but being able to engage in a specific action because we have learnt how to do this action in the past. We may, for instance, remember how to swim but not remember at which specific time and location we learnt how to swim. Nevertheless, our lack of remembering that specific experience in our past does not stop us from being able to swim if we have learnt how to swim in the past (Martin & Deutscher, 1966, 161). I will leave out further considerations regarding remembering-how in this paper as remembering how to do a certain thing usually requires more than one specific event. For instance, when you learnt how to swim, you might have not gotten it on the first stroke but had to try for a few minutes first to stay afloat. Step by step and maybe even on multiple occasions, you have gotten better at swimming and do not just stay afloat but are able to swim quickly from one end of the lake to the other. In this sense, remembering-how might also use memory traces but usually multiple ones at the same time and with different content than in episodic memory as we do not need to remember in which exact way we have learnt to swim in order to be able to swim but only how we have done it (and that can even be done in a subconscious way).

from brain research. Findings suggest that remembering and imagining are linked to a broader range of forms of episodic imagination (Buckner – Carroll 2007; Hassabis & Kumaran & Maguire 2007a; Hassabis & Maguire 2007b, 2009; Schacter & Addis 2007a, 2007b; Schacter et al. 2007c). Episodic imagination could, for example, be imagining the future, the counterfactual future, the present or the counterfactual present, the past or the counterfactual past. Even phenomena such as mind wandering or dreaming may count as episodic imagination. For instance, when we imagine walking around in a familiar city, we also draw on remembered events as our building blocks to represent us walking around in a familiar city. Even when we are just imagining, we use remembered events in order to be able to imagine. Moreover, when we remember, we likewise use our imagination in order to achieve representing an event we have experienced. This motivates Michaelian to introduce simulationism. According to simulationism, episodic imagination includes both remembering and imagining and is produced by a general episodic construction system. This episodic construction system, by producing different forms of episodic information, flexibly transforms and recombines stored information from a variety of sources in order to produce representations of different episodes. Remembering, as it is part of the episodic construction system, is thus a specific form of imagination.

Remembering differs from other forms of episodic imagination when it comes to the target of its representation. In case of remembering, the episodic construction system must aim at imagining an episode from the subject's personal past. It is important to highlight that it is not the subject's target which is of importance here but the aim of the episodic construction system as a person might misclassify their own intentions. They might for instance, think that they are trying to imagine the past while the episodic construction system generates a representation of the future. Thus, it is the 'intention' of the episodic construction system to imagine the past which distinguishes remembering from other forms of imagination. The only difference between remembering the past and imagining the future lies in the target of the episodic construction system and remembering is thus defined as imagining the past. Therefore, the difference between remembering and imagining is a matter of degree (in their target of representation) but not

in kind as both are forms of episodic imagination and these forms of imagination are part of the same mechanism (Michaelian 2016c).

Furthermore, if remembering is one form of episodic imagination, Michaelian (2016a, 104) explains, then our common understanding of remembering does not necessarily apply anymore. While we would usually assume that we need to at least draw on some information originating in the experienced event of our personal past in order to remember and would usually assume that we thus need to have something like a memory trace present in order to ensure that we actually remember, a memory trace becomes superfluous under simulationism (Michaelian 2022, 3). We may still have memory traces, even in the case of imagining in order to draw information from them but they are not a necessary condition for remembering or imagining. As long as we are able to imagine the personal past with the intention of our episodic construction system to imagine the personal past, it does not matter whether we draw on information originating in an experience of the experienced episode of our personal past, but it is only of importance whether we are able to imagine our personal past with our episodic construction system. The information we draw on may originate in an experience of the specific episode but does not necessarily have to. If we are able to imagine the personal past with the intention to imagine the personal past but draw on other information originating in other events of our past, this process is still to be seen as remembering as remembering is to imagine the personal past according to Michaelian and does not come with the requirement to be causally rooted in an event we have actually experienced (Michaelian 2016a.).

3. Empirical evidence defying simulationism

Michaelian claims that remembering and imagining being linked to a broader episodic imagination system implies that they only differ in degree or namely only differ in the intention of the episodic construction system (2016a, 98f.). If remembering and imagining both use the same mechanism and only the outcome of the mechanism is slightly different, we could argue for a difference in degree only. Yet, if remembering and imagining were to differ in degree only, they would need to function equally well. In the

following, I will show that they do not. There are cases in which our ability to imagine remains unimpaired while we are not able to remember. This implies that there is not only a gradual difference between the both of them but a difference in kind. While my main argument will rely on empirical studies which support my claim that the difference between remembering and imagining cannot only be gradual because remembering and imagining do not function equally well, I will also shortly introduce a theory of how this difference in quality could be explained by different brain mechanisms. While I will introduce this possible explanation, the truth of it is not necessary for my argument to be valid. What is necessary is that remembering and imagining do not always function equally well.

In what follows, I will firstly show how a paper Michaelian cites to show that remembering and imagining are the same in kind is also suggestive of there being significant differences between the both of them. Moreover, Michaelian also takes cases of people with amnesia to be suggestive of remembering and imagining being of the same kind. I will secondly outline his claim and subsequently show that research with amnesic people can also be found indicative of remembering and imagining coming apart in kind. Thirdly, I will add a study which suggests that the ability to remember and to imagine comes apart in people without any memory impairment as well.

Michaelian (2016a) suggests that empirical research is indicative of the same brain areas, namely the episodic construction system, being used for remembering and imagining and proposes to see remembering and imagining to be of the same kind due to these empirical results. While it might already be a questionable attempt to decide whether something differs in degree or kind on the basis of neuronal connections, I will take his suggestion at face value here and show how I could oppose his claim. Michaelian (2016a, 98ff.) cites a paper by Schacter and Addis (2007a) to support his claim that remembering and imagining both have the same underlying structure of the episodic construction system. However, while the given paper supports this claim by explaining that there is a considerable overlap of neural and psychological processes involved in imagining and remembering, it also points out significant differences between both. When it comes to recognizing what we have actually experienced and thinking to recognize

something we did not actually experience, different brain activity can be found (Schacter & Addis 2007a, 777) depending on whether we have experienced something or not. Schacter and Addis (2007a, 779) also explain that D'Argembeau and van der Linden found that imagined future events were less vivid when it came to sensory and contextual details than remembered past events. More importantly, though, is their finding of a study by Okuda et al. (1998) which has found that, in contrast to cases of recalling the past, there was a strong positive correlation between imagining the future and right frontopolar activity. Moreover, when activity between constructing past and future events was measured, several brain regions were significantly more active when it came to imagining the future than when it came to remembering the past. These brain areas included the bilateral premotor cortex and left precuneus (Schacter et al. 2007a, 780f.). The introduced counterevidence already provides some room for criticism on Michaelian's theory. However, apart from an explanation of how remembering and imagining work neuronally, there is some counterevidence to simulationism which bites harder.

Michaelian (2016a, 98) proposes additional evidence for remembering and imagining being of the same kind by explaining that Tulving (1985) had already tentatively linked the ability to imagine the future to the ability to remember the past because amnesic patients had been found to be unable to remember past episodes and were also impaired in imagining future episodes. Moreover, further evidence can be found for remembering the personal past and imagining the personal future at least being correlated. People with damage to their hippocampus have been found to have an episodic memory impairment and at the same time an impaired ability to imagine the personal future (Tulving 1985; Klein et al. 2002; Rosenbaum et al. 2005; see Addis et al. 2007 for neuroimaging evidence; Kwan et al. 2010, Juskenaite et al. 2014, De Luca et al. 2017). The conclusion that these results suggest that remembering and imagining are of the same kind seems reasonable in this context and could well be explained by remembering and imagining both relying on the episodic construction system. Yet, if remembering and imagining only differ in degree, they have to function equally well. While this is the case in the by Michaelian introduced studies, there are also multiple cases in which imagining seems to be unimpaired or only

mildly impaired while the ability to remember is severely impaired. I will introduce the relevant studies in the following.

Studies (Allen, 2018; Andelman, Hoofien, Goldberg, Aizenstein & Neufeld, 2010; Klein, Loftus & Kihlstrom, 2002; Tulving, 1985; Mullally, S. L. & Maguire, E. A., 2014) suggest that patients with amnesia can imagine events which have happened to other people without severe limitations compared to healthy control subjects while their ability to remember their past is severely impaired. Moreover, people with memory impairments were still found to be able to imagine future experiences and people's possible pasts, people's real pasts and presents in a slightly impoverished fashion while their ability to remember their own past was highly impaired. Juskenaite et al. (2014), for instance, have examined people with transient global amnesia[6]. They found that these people were able to imagine personal future events nearly as well as healthy control participants if there was a short description of common scenarios present. However, they struggled with recalling personal past events when they were asked to do so. They produced significantly fewer past events than control groups and the events were scant and contained less details. Cooper, Vargha-Khadem, Gadian, and Maguire (2011), have reported that 21 children with developmental amnesia[7] displayed an unimpaired ability to imagine new events when a short description of possible events such as "Imagine you are by a campfire in the mountains" was given to them. The imagined events were similar to events children have already experienced multiple times (such as sitting by a campfire). At the same time, the same people could not remember their own personal past or only to a very limited extent. Maguire, Vargha-Khadem

[6] Transient global amnesia (TGA) is a condition characterized by sudden onset of memory loss and confusion. During an episode of TGA, a person is not able to make new memories. The person may be disoriented in regard to time and place, but can remember who they are and can recognize family members. TGA typically lasts for several hours, but can last up to ten hours. Since no memories are made during a TGA episode, the person will never remember what happened during this period, but all other memory is usually intact. Most people have only one episode of TGA during their lifetime. The underlying cause of TGA is unclear (Genetic and Rare Diseases Information Center 2021).

[7] Developmental amnesia is amnesia with an early onset. Namely, when neural plasticity is at its peak in children (Vargha-Khadem et al. 2003).

and Hassabis (2010) have shown that another patient's ability to imagine new events which they had experienced multiple times, when supplied with a short description of these experienced events, was intact but not their ability to imagine old, experienced events. Hurley, Maguire, and Vargha-Khadem (2011) have reported similar results on the same task for a person suffering from developmental amnesia (Kwan, Carson, Addis & Rosenbaum, 2010). Rosenbaum et al. (2009) have examined a patient with severe anterograde and retrograde amnesia. Retrograde amnesia is the inability to remember events before the onset of amnesia while anterograde amnesia is the inability to remember events after the onset of amnesia. Retrograde and anterograde amnesia can also appear simultaneously (Gilboa et al. 2006). The person Rosenbaum et al. (2009) had examined was able to reconstruct semantic information of their past and future and was not only able to come up with narratives but was also able to distinguish between non-personal semantic narratives and personal ones in the settings of the experiment even though they were not able to in episodic narratives. They were able to generate fictional events (such as, what people will do at a birthday party) and to recall and recognize details of well-known fairy tales and bible stories to a lesser extent than people without this impairment but still too a high degree. Their recall was more skeletal and gist-like than in healthy control participants, but the patient was able to come up with a consistent story. Yet, the same patient was also described as follows: he has a high number of semantic but a low number of episodic or autobiographic memories. His semantic and procedural memory seem to be unimpaired, but his episodic memory is highly impaired. He cannot experience extended subjective time. He feels like he has a personal identity, but this does not extend to past or future. He seems to be living in a permanent present (Tulving 1985). Cermak and O'Connor (1983) report on a person with severe anterograde and retrograde amnesia caused by a case of encephalitis. Despite their severe case of amnesia, they could still encode information at a semantic level and were able to rely on semantic information when it came to generating events of what could have happened in their past. Yet, they were not able to recall episodes of their lives, neither personal nor public ones. O'Connor et al. (1992) also report that a young amnesic person they had worked with was still able to recall factual, semantic information of the past with some

impairments but had severe problems recalling personal episodic events. In this severe case, the person was not able to recall any events before the onset of their illness including their high school years and their early childhood or the fact that their parents had been divorced three years before the onset of their amnesia, names of their childhood friends or high school graduation.

The introduced studies suggests that people with amnesia can still imagine non-personal episodes such as the past, the future, the possible past and the possible future and also their own potential future if they are provided with a short description about what they had already experienced in their past. However, at the same time, they are not able to reconstruct their own personal actual past to an equal qualitative and quantitative level. Their ability to imagine and their ability to remember comes apart. If remembering and imagining were of the same kind, this should not be the case as an impairment at one level should have the same effects on all kinds of episodic imagination. Therefore, I propose that remembering and imagining are different in kind.[8]

Moreover, evidence for a significant difference between remembering and imagining cannot only be found in people with amnesia but also in people without any neural impairments. The first, and at that time only, lifespan study about episodic and semantic past, present and future autobiographic memory of all age groups from childhood to older adulthood suggests that remembering and imagining could have a fundamental different basis. This claim is based on the finding that imagining the personal future remains

[8] Some of the introduced studies seem to suggest that remembering and imagining may not be so different after all as imagining the past and remembering seem to be at an equal level of impairment in amnesic people. Firstly, I want to remark that it is hard to distinguish between whether a person imagines or remembers the past in an empirical experiment. Yet, even if I should grant the point of remembering and imagining the past being equally impaired here, the studies still support my distinction in a quantitative fashion as people with amnesia can still imagine non-personal episodes such as the past, the future, the possible past and the possible future and also their own potential future if they are provided with a short description about what they had already experienced in their past, but, at the same time, are not able to reconstruct their own personal actual past to an equal qualitative and quantitative level.

equally difficult for all age groups even though their episodic memory is at different levels of quality depending on people's age. 6-8-year-old children were found to be truly unable to imagine or remember personal events. Because of the parallel inability to remember and imagine, one could initially assume that remembering and imagining are part of the same mechanism (Abram et al. 2014). However, Naito and Suzuki's (2011) study suggest that future episodic autobiographic memory abilities develop after the age of five while the ability to episodically remember is developed to a fuller extent at the age of four already. Up until young adulthood, the ability to episodically remember increases while a general decline in this ability can be found after young adulthood. The ability to episodically remember can be illustrated in an upside-down turned U-formed graph. While the lowest points would be early childhood and older adulthood, the highest point would be young adulthood. With decreasing episodic remembering ability, semantic remembering fills in and is more commonly used in older adults than in younger ones[9]. As there are different abilities to episodically remember due to a U-shaped development in people, one should assume that the ability to episodically imagine the future would also decline in age and grow from childhood up until young adulthood. However, except for 6–8-year-old children, in which both abilities were not clearly given yet, the ability to episodically imagine the future has been found to remain at an equal level independent of age groups (Abram et al. 2014).[10]

This study shows a difference between remembering and imagining. While imagining seems to stay at an equal level of ability independent of

[9] Addis, Schachter and Roberts (2011) have also proposed that imagining and remembering underlie the same constraints when it comes to episodic and semantic memory. Older adults have been found to not only draw from more semantic information when it comes to remembering but also when it comes to imagining. This, however, is not counterevidence to my claim as the ability to imagine, according to Abram et al., 2014, stays at an equal level while the ability to remember declines. For my argument, it does not matter how imagining or remembering is achieved but whether it is at the same quantitative level.

[10] There are studies which suggest that the ability to imagine one's own future also declines with age (i.e. Than 2008). Yet, the given studies mainly describe a switch from episodic to semantic information, not, however, a decline in the ability to imagine.

age, the ability to remember is fluctuating. Yet, if remembering and imagining were of the same kind, this should not be the case. Once people have problems to access their personal past, they should also have problems in imagining their personal future and an increased ability in accessing one's memory should cause people to be able to imagine their personal future better than with less ability to access their memory.

The cited studies suggest that there is a difference in the ability to remember one's past and to imagine in people with amnesia but also in people with non-impaired memory. If remembering and imagining were of the same kind, this should not be the case. If they were of the same kind, they should be equally impaired. I also assume that this difference cannot be explained by a difference in intention in the episodic construction system. Having a different intention should not cause a gap in what we are able to remember and imagine. Therefore, I suggest that they come apart in more than just the intention of the episodic construction system. Remembering and imagining are of different kind.

4. A Causal theory of memory

In the last chapter, I have introduced empirical studies which defy the claim of simulationism that remembering and imagining are of the same kind. While they both may rely on the same system, they seem to need additional processes to actually constitute remembering or imagining. Therefore, I claim that simulationism cannot be right. Yet, it's opposing view of the causal theory of memory also cannot explain the difference between remembering and imagining. The original distinction consisted in remembering needing a memory trace. However, in order to imagine, we also need to imagine from something and this something we imagine from is usually also something we have experienced. Thus, remembering and imagining are both in need of memory traces. In what follows, I will suggest that a slightly changed version of the causal theory of memory enables us to make a distinction between remembering and imagining which is consistent with the introduced research data. I will introduce the slightly changed version first and will subsequently explain how it helps us in making a distinction between remembering and imagining.

Martin and Deutscher did not describe how a memory trace would look like from a neurological standpoint of view. However, there is empirical evidence for memory traces in the form of brain reactivation. According to the so-called sensory reactivation theory, there is reactivation of the same brain areas present which were active while experienced events got encoded (Rugg et al. 2008). Additionally, the sensory reactivation theory can also be supported by the fact that memory does not come with one dedicated encoding circuit. Depending on which information gets encoded, different brain areas are active (Rugg et al. 2008). Therefore, we should have reactivation of the same brain areas which were active in encoding during recall. This seems to be the case. Addis et al (2009) explain that neuroimaging studies of memory for previously studied pictures have revealed reactivation during retrieval of some of the same visual processing regions that were active during encoding. Woodruff et al. (2005) have found a dissociation in activation for brain areas when it came to recall of pictures and words. So, a different brain area was active when a word was recalled than when a picture was recalled. Kahn et al. (2004) found that tasks which require either visual imagery or phonological processing elicited a greater response in their respective brain areas during retrieval. So, words from the phonological condition elicited greater activity in areas responsible for phonological processing while words from visual imagery elicited greater activity in its respective brain areas during recall. Rugg et al. (2008) found that people who were asked to study words in four different locations and colors were found to have higher brain activity in, among other regions, their retrosplenial cortex if they correctly reported the location of the word and higher activity in their posterior inferior temporal cortex if they correctly recalled the color of the word. The retrosplenial cortex has been previously found to be active in the process of encoding information when it came to location and the posterior inferior temporal cortex when it came to encoding information about color (Chao and Martin 1999; Kellenbach et al. 2001; Mayes et al. 2004; Frings et al. 2006). The sensory reactivation theory also leaves room for the possibility that we do not recall something we have experienced one to one but only with deviations. The sensory reactivation theory does not demand a perfect overlap between an originally encoded event and brain activity during retrieval. Johnson and Rugg (2007), for instance, report a

limited overlap between brain activity in encoded information and retrieved information. Yet, Rugg et al. (2008) suggest that even activity which only partially overlaps with formerly encoded information may cause the reactivation of the whole former representation. This can be explained by memories usually being distorted and being invariably partial records of an originally experienced event (Bartlett 1932; Loftus & Palmer 1974, Schacter 2002). With the sensory reactivation theory, we have a possible scientific explanation of what memory traces may look like. In the following, I will also explain how this theory helps the causal theory of memory to make a distinction between remembering and imagining.

According to Martin and Deutscher (1966), we remember if a memory trace is operative in recalling an experienced event. If a memory trace is not operative in recalling an experienced even, we do not remember. Applied to the sensory reactivation theory, this would mean that we remember if there is the same reactivation of formerly active brain areas present in recall than was present while we have experienced the event we recall. However, the mere presence of a memory trace is also not sufficient to distinguish between remembering and imagining because imagining could in principle also come with a memory trace. After all, we need to have some input to imagine even a counterfactual event from and this might as well be done with the help of a memory trace. According to Michaelian (2016a, 99 ff.), this way of reasoning has led brain research to the suggestion that remembering and imagining are linked to a broader range of forms of episodic imagination and which had led him to propose the gradual distinction between remembering and imagining. The possibility that we use memory traces in imagining cannot be excluded and I will not argue against it. Memory traces may as well be active in the case of imagination. I claim, that remembering and imagination, cannot be distinguished by the presence of a memory trace but by whether a memory trace is solely operative in representing an experienced event. Schacter and Addis (2007a) suggest that remembering and imagining both need information from memory. However, only imagination requires flexible recombination of these events in order to be able to recombine the given events into a new event. Remembering, by just representing the past, does not need recombination. Thus, even if there are memory traces involved in representing an event, it is not the sole

responsibility of the memory trace to produce an imagination. In order to imagine, we do not only need memory traces but further processes on top of memory traces. In remembering, however, these processes of recombination are not needed. It suffices if the memory trace is operative in producing a representation of a past event for us to remember. Therefore, I claim that the distinction between remembering and imagining can be drawn by answering the question of whether a memory trace was solely operative in producing a representation. If a memory trace was solely operative in producing a representation, we remember. If we need more than the operative presence of a memory trace, we imagine. This claim can also be supported by Addis et al. (2009). They suggest that remembering and imagining both need to draw information from memory but that they are divided into two subsystems. Remembering past events is associated with greater recruitment of a remembering subsystem than is imagining events. This subsystem included posterior visual cortices, such as fusiform, lingual and occipital gyri and cuneus, in addition to regions previously associated with remembering past events (i.e., medial prefrontal, hippocampus and parahippocampal gyrus) while an imagining subsystem network included extensive aspects of bilateral medial prefrontal cortex, inferior frontal gyrus, medial temporal lobe, polar and posterior temporal cortex, medial parietal cortex and cerebellum.

In this paper, I have argued against the claim of simulationism that remembering and imagining are of the same kind. I have cited studies that suggest that simulationism is defeated. Moreover, I have also given a possible explanation of how we can actually make a distinction between remembering and imagining. Both remain separate entities

References

Abram, M., Picard, L., Navarro, B., Piolino, P. 2014. "Mechanisms of Remembering the Past and Imagining the Future: New Data from Autobiographical Memory Tasks in a Lifespan Approach." *Consciousness and Cognition*, 29: 76–89. https://doi.org/10.1016/j.concog.2014.07.011

Addis, Dana Rose, Wong, Alana T., Schacter, Daniel L. 2007. "Remembering the Past and Imaging the Future: Common and Distinct Neural Substrates during Event Construction and Elaboration." *Neuropsychologia*, 45: 1363–77. https://doi.org/10.1016/j.neuropsychologia.2006.10.016

Addis, Dana Rose; Pan, Ling; Vu, Mai Ahn; Laiser, Noa; Schacter, D.L. 2009. "Constructive Episodic Simulation of the Future and the Past: Distinct Subsystems of a Core Brain Network Mediate Imagining and Remembering." *Neuropsychologia*, 47: 2222-38. https://doi.org/10.1016/j.neuropsychologia.2008.10.026

Allen, Richard.J. (2018). "Classic and Recent Advances in Understanding Amnesia." *F1000Research*. 10.12688/f1000research.13737.1

Andelman, Fani; Hoofien, Dan; Goldberg, Iian; Aizenstein, Orna; Neufeld, Miri. 2010. "Bilateral Hippocampal Lesion and a Selective Impairment of the Ability for Mental Time Travel." *Neurocase*, 16: 426–35.

Bartlett, Fredireck. 1932. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: Cambridge University Press.

Buckner, Randy; Carroll, Daniel. 2007. "Self-Projection and the Brain." *Trends in Cognitive Sciences*, 11(2): 49–57. https://doi.org/10.1016/j.tics.2006.11.004

Cermak, L.S., O'Connor, M. 1983. "The Anterograde and Retrograde Retrieval Ability of a Patient With Amnesia Due to Encephalitis." *Neuropsychologia*, 21(3): 213-34. https://doi.org/10.1016/0028-3932(83)90039-8

Chao, L.L., Martin, A. 1999. "Cortical Regions Associated with Perceiving, Naming, and Knowing About Colors." *Journal of Cognitive Neuroscience*, 11: 25–35. https://doi.org/10.1162/089892999563229

Cooper, Janine; Vargha-Khadem, Faradeh; Gadian, David; Maguire, Eleanor. 2011. "The Effect of Hippocampal Damage in Children on Recalling the Past and Imagining New Experiences." *Neuropsychologia*, 49(7): 1843–50. https://dx.doi.org/10.1016/j.neuropsychologia.2011.03.008

De Luca, Flavia; Benuzzi, Franchesca; Bertossi, Elena; Braghittoni, Davide; di Pellegrino, Giuseppe; Ciaramelli, Elisa, 2017. "Episodic Future Thinking and Future-Based Decision-Making in a Case of Retrograde Amnesia." *Neuropsychologia*, 110: 92–103. https://doi.org/10.1016/j.neuropsychologia.2017.08.007

Frings, Lars; Wagner, Kathrin; Quiske, Ansgar; Schwarzwald, Ralf; Spreer, Joachim; Halsband, Ulrike and Schulze-Bonhage, Anreas. 2006. "Precuneus Is Involved in Allocentric Spatial Location Encoding and Recognition." *Exp. Brain Res.*, 173: 661–72. https://doi.org/10.1007/s00221-006-0408-8

Gilboa, Asaf; Winocur, Gordon; Rosenbaum, Shayma; Poreh, Amir; Gao, F., Black, S. E., Westmacott, Robyn; Moscovitch, Morris. 2006. "Hippocampal Contributions to Recollection in Retrograde and Anterograde Amnesia." *Hippocampus*, 16: 966–80. https://doi.org/10.1002/hipo.20226

Hassabis, D., Kumaran, D., Vann, S.D., Maguire, E.A. 2007a. "Patients with Hippocampal Amnesia Cannot Imagine New Experiences." *Proceedings of the National Academy of Sciences of the United States of America*, 104(5): 1726–31.

Hassabis, Demis; Maguire, Eleamor. 2007b. "Deconstructing Episodic Memory with Construction." *Trends in Cognitive Sciences*, 11(7): 299–306.

Hurley, Niamh; Maguire, Eleanor; Vargha-Khadem, F. 2011. "Patient HC with Developmental Amnesia Can Construct Future Scenarios." *Neuropsychologia*, 49: 3620–28.

Johnson, J. D., Rugg, M.D. 2007. "Recollection and the Reinstatement of Encoding Related Cortical Activity." *Cerebral Cortex*, 17: 2507-15.

Juskenaite, Aurelija; Quinette, Peggy; Desgranges Beatrice; de La Sayette, Vincent; Viader, F., Eustache, F. 2014. "Mental Simulation of Future Scenarios in Transient Global Amnesia." *Neuropsychologia*, 63: 1-9. https://doi.org/10.1016/j.neuropsychologia.2014.08.002

Kahn I, Davachi L., Wagner A.D. 2004. "Functional-Neuroanatomic Correlates of Recollection: Implications for Models of Recognition Memory." *Journal of Neuroscience*, 28: 4172–80.

Kellenbach, Marrion; Brett, Matthew and Patterson, Karalyn. 2001. "Large, Colorful, Or Noisy? Attribute-and Modality-Specific Activations during Retrieval of Perceptual Attribute Knowledge." *Cogn. Affect. Behav. Neurosci.*, 1: 207–21. https://link.springer.com/article/10.3758/CABN.1.3.207

Klein Stanley; Loftus Judith; Kihlstrom John. 2002. "Memory and Temporal Experience: The Effects of Episodic Memory Loss on an Amnesic Patient's Ability to Remember the Past and Imagine the Future." *Social Cognition*, 20: 353–79. https://doi.org/10.1521/soco.20.5.353.21125

Kwan D, Carson N, Addis D.R., Rosenbaum R.S. 2010. "Deficits in Past Remembering Extend to Future Imagining in a Case of Developmental Amnesia." *Neuropsychologia*, 48: 3179–86.

Loftus, Elizabeth; Palmer, John. 1974. "Reconstruction of Automobile Destruction." *J. Verbal Learn. Verbal Behav.*, 13: 585–9.

Maguire, Eleanor; Vargha-Khadem, Faraneh; Hassabis, Demis. 2010. "Imagining Fictitious and Future Experiences: Evidence from Developmental Amnesia." *Neuropsychologia*, 48: 3187–92.

Martin, C.B., Deutscher, M. 1966. "Remembering." *The Philosophical Review*, 75(2): 161-96.

Mayes, A.R., Montaldi, D., Spencer, T.J. and Roberts, N. 2004. "Recalling Spatial Information as a Component of Recently and Remotely Acquired Episodic Or Semantic Memories: An Fmri Study." *Neuropsychology*, 18: 426–41.

Michaelian, Kourken. 2016a. "Mental Time Travel Episodic Memory and Our Knowledge of the Personal Past." In *Life and Mind: Philosophical Issues in Biology and Psychology*, ed. by K. Sterelny, R.A. Wilson. Cambridge, Massachusetts: MIT Press.

Michaelian, Kourken. 2016b. "Confabulating, Misremembering, Relearning: The Simulationist Theory of Memory and Unsuccessful Remembering." *Frontiers in Psychology*, 7: 1857.

Michaelian, Kourken. 2016c. "Against Discontinuism: Mental Time Travel and Our Knowledge of Past and Future Events." *Seeing The Future: Theoretical Perspectives on Future-Oriented Mental Time Travel,* edited by K. Michaelian, S.B. Klein, S.B., K.K. Szpunar, 62-92. Oxford: Oxford University Press.

Michaelian, Kourken., Robins, Sarah. 2018. "Beyond the Causal Theory? Fifty Years after Martin and Deutscher." *New Directions in the Philosophy of Memory*, edited by K. Michaelian, D. Debus, D. Perrin, 13-32. London and New York: Routledge.

Michaelian, Kourken. 2022. "Radicalizing Simulationism: Remembering as Imagining the (Nonpersonal) Past." *Philosophical Psychology.* https://doi.org/10.1080/09515089.2022.2082934

Mullally, Sinead; Maguire, Eleanor. 2014. "Memory, Imagination, and Predicting the Future: A Common Brain Mechanism?" *The Neuroscientist* 20(3), 220–34.

Naito, Mika; & Suzuki, Toshiko. 2011. "When Did I Learn and When Shall I Act?: The Developmental Relationship between Episodic Future Thinking and Memory." *Journal of Experimental Child Psychology*, 109, 397–411. https://doi.org/10.1016/j.jecp.2011.03.005

O'Connor, Margaret; Butters, Nelson; Miliotis, Patti; Eslinger, Paul; Cermak, Laird. 1992. "The Dissociation of Anterograde and Retrograde Amnesia in a Patient with Herpes Encephalitis." *Journal of Clinical and Experimental Neuropsychology*, 14(2): 159–78. https://doi.org/10.1080/01688639208402821

Okuda, Jiro; Fujii, Toshikatsu; Yamadori, Atsushi; Kawashima, R., Tsukiura, T., Fukatsu, R., Suzuki, K., Itoh, M., Fukuda, H. 1998. "Participation of the Prefrontal Cortices in Prospective Memory: Evidence From a PET Study in Humans." *Neurosci. Lett.* 253: 127–30. https://doi.org/10.1016/S0304-3940(98)00628-4

Rosenbaum, Shayna; Kohler, Stefan; Schacter, Daniel; Moscovitch, Morris; Westmacott, Robyn; Black, Sandra;., Gao, Faqiang; Tulving, Endel. 2005. "The Case of K.C.: Contributions of a Memory-Impaired Person to Memory Theory." *Neuropsychologia*, 43: 989–1021.

Rugg, Michael.D.; Johnson, Jeffrey.D.; Park, Heekyeong; Uncapher, M.R. 2008. "Encoding-Retrieval Overlap in Human Episodic Memory: A Functional Neuroimaging Perspective." *Progress in Brain Research*, 169: 339-352.

Schacter, Daniel. 2002. *The Seven Sins of Memory: How the Mind Forgets and Remembers*. Boston: Houghton Mifflin.

Schacter, Daniel., Addis, D.R. 2007a. "The Cognitive Neuroscience of Constructive Memory: Remembering the Past and Imagining The Future." *Philosophical*

Transactions of the Royal Society of London. Series B, Biological Sciences, 362(1481): 773–86. https://doi.org/10.1098/rstb.2007.2087

Schacter, Daniel; Addis, Dona. 2007b. "Constructive Memory: The Ghosts of Past and Future." *Nature*, 445 (7123): 27.

Schacter, Daniel; Addis, Dona; Buckner, R.L. 2007c. "Remembering the Past to Imagine the Future: The Prospective Brain." *Nature Reviews. Neuroscience*, 8(9): 657-61. https://doi.org/10.1038/nrn2213

Schacter, Daniel; Addis, D.R., Roberts, R.P. 2011. "Age-Related Neural Changes in Autobiographical Remembering and Imagining." *Neuropsychologia*, 49: 3656-69.

Than, K. 2008. "Ageing Makes the Imagination Wither." *Nature*, https://www.nature.com/articles/news.2008.408 (last accessed: 08.12.2023).

Tulving E. 1985. "Memory and Consciousness." *Canadian Psychology*, 26: 1–12.

Woodruff C.C., Johnson J.D., Uncapher M.R., Rugg, M.D. 2005. "Content-Specificity of the Neural Correlates of Recollection." *Neuropsychologia*, 43:1022–32.

Organon F 30 (4) 2023: 372–411
https://doi.org/10.31577/orgf.2023.30405

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

RESEARCH ARTICLE

The Riddle of Understanding Nonsense

Krystian Bogucki*

Received: 4 May 2023 / Revised: 13 December 2023 / Accepted: 16 December 2023

Abstract: Typically, if I understand a sentence, then it expresses a proposition that I entertain. Nonsensical sentences don't express propositions, but there are contexts in which we talk about understanding nonsensical sentences. For example, we accept various kinds of semantically defective sentences in fiction, philosophy, and everyday life. Furthermore, it is a standard assumption that if a sentence is nonsensical, then it makes no sense to say that it implies anything or is implied by other sentences. Semantically uninterpreted sentences don't have logical characteristics. Hence, the riddle of understanding nonsense arises. We seem to use nonsensical sentences in reasoning, thinking, judging, and drawing conclusions, but they convey no propositions, which are the vehicles of their semantic properties. In this article, I propose the pretence theory of understanding nonsense to explain the riddle of understanding nonsense, and discuss alternative frameworks that are insufficient to solve it.

Keywords: Nonsense; pretence; understanding, fiction, category mistakes.

* Polish Academy of Sciences
https://orcid.org/0000-0002-4753-497X
Institute of Philosophy and Sociology, Polish Academy of Sciences, Nowy Świat 72, 00-330 Warsaw, Poland.
krystian.bogucki@ifispan.edu.pl

1. Introduction

The riddle of understanding nonsense is the following set of problems. Typically, if I understand a sentence *s*, then I entertain proposition *p*, which is expressed by the sentence *s*. Alternatively, if I understand a sentence *s*, then there are some truth-conditions of this sentence that I understand, or there is a mental representation of the proposition *p* that is the content of my belief.[1] In each case, the sentence *s* expresses the shared content that is responsible for its understanding. Nonsensical sentences, on the other hand, don't express propositions (or truth-conditions if you prefer this semantic framework). Nonsensical sentences have no meaning; hence, they don't express propositions or truth-conditions. Furthermore, it is a standard assumption that if a sentence is nonsensical, then it makes no sense to say that it implies anything or is implied by other sentences (Glock 2004, White 2011, Whiterspoon 2000, cf. McManus 2014). Logical relations occur between contents of sentences: some propositions imply other propositions and are implied by another ones. Uninterpreted sentences don't possess logical characteristics. The contents of nonsensical sentences cannot stand in logical relations since nonsensical sentences don't have contents (or, e.g., truth-conditions).

In this article, I propose a pretence theory of understanding nonsense. A theory of understanding nonsense is a theory that aims to explain how we understand nonsensical sentences, how we can draw consequences from nonsensical sentences, how we can reason with nonsensical sentences, and how our understanding of nonsensical sentences differs from our understanding of meaningful sentences. In short, the pretence theory of understanding nonsense is supposed to indicate the mechanism behind understanding nonsensical sentences, and it is a solution to the riddle of understanding nonsense. The concept of pretence has been used to explain a wide range of phenomena: there are pretence accounts of fiction (e.g., Walton

[1] The details of these stories depend on the preferred theory of language: whether it should be expressed in terms of propositions, truth-conditions, the language of thought, *etc.* The story about what meaningful sentences express can also be formulated in terms of use. These reservations will be taken into account in the later parts of the article.

1990, 2015), existence (e.g., Evans 1982, Walton 1990), truth and reference (e.g., Armour-Garb and Woodbridge 2013, 2015), mathematics (e.g., Armour-Garb and Woodbridge 2015, Yablo 2001), semantics of attitude ascriptions (Crimmins 1998) and many others.[2] The notion of pretence is also used to account for some specific features of other kinds of discourse; for instance, most theories of fiction appeal to it in some way (e.g., Lewis 1978, Searle 1975, Thomasson 1999, 2003).[3] So, it has already been proved that the notion of pretence is fruitful, and I believe that the notion of pretence is the correct solution to the riddle of understanding nonsense.

The paper is organised as follows. In Section 2, I will show why the riddle of understanding nonsense is important. In Section 3, I will present a framework that is the correct answer to the riddle of understanding nonsense. In Section 4, I will point out the shortcomings of Manish Oza's (2022) pretence account of nonsense. In Section 5, I will show that L. J. Keller and J. A. Keller's (2021) and Recanati's (1997) positions, which attempt to solve the riddle of understanding nonsense by means the language of thought hypothesis, are not satisfactory.

2. The significance of understanding nonsense

In my view, there are some important domains in which the notion of understanding nonsense is useful and widespread. The first such domain is fiction of all kinds, especially fairy tales and children's literature.[4] Fairy

[2] These are only representative works and the list isn't complete. For an interesting discussion and classification, see Armour-Garb and Woodbridge (2015, Ch. 1).

[3] For a discussion of this thesis, see again Armour-Garb and Woodbridge (2015, 1.5.2). Garcia-Carpintero (2019) provides a useful discussion of the landscape of fiction theories.

[4] In this section I want to be as neutral as possible towards the various theories of nonsense. I propose discussing different areas where the idea of dealing with nonsensical sentences seems to play a role. I don't want to propose any unification of these examples, leaving that to theorists of nonsense. I start with category mistakes because they seem to me to be emblematic cases of nonsense based on historical work by Carnap (1959), Ryle (1949, 2009) and Russell (1908, 1910). It is possible

tales commonly describe magic events and unusual characters; many of their sentences involve category mistakes.[5] In a fairy tale, for instance, a pot can talk, see and hear. It can drink coffee and talk to other characters.[6] There are no problems with these actions in the imagined world of fiction; however, from the point of view of the most prominent theory of nonsense, sentences describing these actions are problematic. Let's note the sentence

(1) The pot is drinking coffee.

This sentence is nonsensical because it attributes to an inanimate thing a property that can only be ascribed to other types of stuff (animate objects). Only living things can breathe, eat and drink. Of inanimate objects, it might make sense to say that they are heavy, pretty or ceramic. Structurally (1) is similar to (2) and (3):

(2) Julius Caesar is a prime number.

---

that recent work on category mistakes has made people more critical of what I call below the standard view, which postulates that category mistakes are nonsensical (Camp 2004, Magidor 2009, 2013).

[5] The term "category mistakes" is ambiguous. On the one hand, it can denote a class of sentences that are semantically or pragmatically infelicitous. This way of speaking (let's call it "semantical") just identifies an appropriate type of sentence in order to pose the problem of the nature of its infelicity. Ofra Magidor, for instance, uses the term "category mistakes" in this way in the title of her paper "Category Mistakes are Meaningful". In her book, "Category Mistakes", she does not define the titular notion but introduces it through examples (2013, 1). This way of speaking (let's call it "logical") stipulates that category mistakes are nonsensical because of a violation of logical syntax. I use the term "category mistake" in the former sense. This way of using the term "category mistake" is neutral towards different theories of category mistakes. This is a very important feature since it makes it possible to talk about sentences like "Julius Caesar is a prime number" without siding with any conception of nonsense. Proponents of the no nonsense view, such as Magidor, claim that category mistakes are semantically meaningful (though pragmatically infelicitous). Adherents of the austere approach claim that category mistakes are nonsensical, but their account of the source of nonsensicality differs from the standard view of nonsense.

[6] This example has been discussed by Ludwig Wittgenstein in *Philosophical Investigations* (§281). For a discussion, see Glock (2004, 241).

(3) The theory of relativity is eating breakfast.

Each of these sentences contains some sort of category mistake, and according to the standard view of nonsense these category mistakes are violations of the rules of logical syntax. The standard view of nonsense holds that the main source of nonsense is sentences that violate the rules of logical syntax.[7] Some sentences violate the rules of logical syntax because their components cannot be connected on the basis of logical categories of words.[8] The standard view postulates the division of words into logical categories (e.g., material objects, properties of material objects, properties of properties of material objects, etc.; abstract objects, their properties, properties of these properties, animate objects and their properties, etc.). Sentences containing words with incompatible logical categories of this sort are nonsensical. (1), (2) and (3) violate the rules of logical syntax and are nonsensical. If one asserts (2), one makes a category mistake of attributing the property of being an abstract object to a material object. If one asserts (3), one makes a category mistake of ascribing to an abstract object a property enjoyed only by animate objects. In fact, sentence (2) is Rudolf Carnap's famous example illustrating how sentences that violate the rules of logical syntax result in category mistakes and thus in nonsensical statements (Carnap, 1959). Carnap says in *The Elimination of Metaphysics Through Logical Analysis of Language*:

---

[7] The most prominent defenders of the standard view are Russell (1908, 1910), Carnap (1959) and Ryle (1938, 1949). More recent defences of it can be found in Hacker (2003) and Glock (2015). Some Wittgensteinians call this approach to nonsense "the substantial view of nonsense" because it distinguishes between "mere" nonsense (e.g., "John is xwwwy"), which contains a component without meaning, and a more substantial kind of nonsense (e.g., "Julius Caesar is a prime number"), which requires a notion of logical syntax. Carnap (1959) and Hacker (2003) also give a fairly clear description of what logical syntax is.

[8] Ryle (2009, 178) characterises category mistakes as follows:
When a sentence is (not true or false but) nonsensical or absurd, although its vocabulary is conventional and its grammatical construction is regular, we say that it is absurd because at least one ingredient expression in it is not of the right type to be coupled or to be coupled in that way with the other ingredient expression or expressions in it. Such sentences, we may say, commit type-trespasses or break type-rules.

> Another very frequent violation of logical syntax is the so-called "type confusion" of concepts. (...) An artificial example is the sentence we discussed earlier: "Caesar is a prime number". Names of persons and names of numbers belong to different logical types, and so do accordingly predicates of persons (e.g., "general") and predicates of numbers ("prime number") (Carnap 1959, 75).

Fiction offers much more nonsense than the aforementioned category mistakes. It is common in Wittgensteinian literature to discuss various uses of nonsense in Lewis Carroll's *Alice in Wonderland* books (Glock 2004, 2015, White 2011, McManus 2014). These literary examples include nonsense poems such as *Jabberwocky*, nonsensical uses of the word 'nobody', and metaphysically impossible events, such as the appearance of the Cheshire Cat. We seem to understand what is going on in all these passages, even though they violate the rules of language. *Jabberwocky* consists of nonsensical words invented by Carroll. The uses of the word 'nobody' violate the standard syntactic role of 'nobody'. The word 'nobody' is used as if it were a proper name rather than an indefinite pronoun. The case of the Cheshire Cat is slightly different as it belongs to the same group as Mauritius C. Escher's works of art (Hacker 2003, White 2011, McManus 2014), which present metaphysically, logically, or mathematically impossible events and produce visual illusions in the viewer.[9] Furthermore, some other techniques used in fiction, such as tropes of personification (e.g., abstract entities like death can ride horses or have hands) or metafictional techniques (e.g., fictitious protagonists can address the audience or 'chase' the author of the book), also constitute exemplary cases of category mistakes (Nolan 2021). Again, according to the standard view of nonsense, category mistakes are primary examples of nonsense since they violate the rules of logical syntax. Yet, it is plausible that we somehow understand these nonsensical stories.

[9] One can question whether these are real instances of nonsense. At least some Wittgenstein scholars would vehemently argue that they are. For the record, Glock (2004, 239) and White (2011) provide other literary and non-literary cases of understanding nonsense that seem far more controversial to me. However, this shows that the class of potential candidates is quite wide.

Every reader of *Tractatus Logico-Philosohpicus* should address the issue of understanding nonsense. Wittgenstein famously claims that many/all sentences (*Sätze*) in his works are nonsensical (*unsinnig*) (§6.54):

> My sentences [Sätze] serve as elucidations in the following way: anyone who understands me eventually recognizes them as nonsensical [unsinnig], when he has used them—as steps—to climb up beyond them. (He must, so to speak, throw away the ladder after he has climbed up it.)
>
> He must overcome [überwinden] these propositions, and then he will see the world aright.

Wittgenstein also gives some specific examples of sentences that he says are nonsensical, such as 'There are objects' and '1 is a number' (§4.1272). However, we can set aside these two controversial instances and focus on another Tractarian example:

(4) A picture is a fact.

In fact, Wittgenstein claims that it is not possible to say anything about formal concepts such as complexes, facts, functions, numbers and pictures without falling into nonsense. *Tractatus* is full of statements of this kind (§4.1272). It very often says things like "a fact is such-and-such", "a function is such-and-such". The conclusion that the *Tractatus* is nonsensical is justified in its own terms. For this reason, the riddle of understanding nonsense arises; the reader thinks she understands Tractarian sentences and draw some conclusions from Tractarian arguments. If Tractarian sentences are nonsensical, then there are no propositions that the reader can understand. If no propositions are expressed by Tractarian sentences, then there are no logical inferences that the reader can entertain. Nonsensical sentences do not possess the logical characteristics which meaningful sentences enjoy, but, at first sight, Tractarian sentences aren't psychologically distinguishable from the latter. It seems inexplicable why we seem to understand Tractarian sentences, why we read them like ordinary, meaningful sentences, and how they can exert any influence on readers. Furthermore, as noted by Peter Sullivan (2003), the influence of *Tractatus* should be conceptual, not causal: it should persuade by means of concepts. The influence on readers of *Tractatus* is not 'like a blow on the head'

(Sullivan 2003). The engagement and appreciation of Tractarian nonsense is conceptual in nature.

The issue of understanding Tractarian sentences may seem to be just a historical problem or a curiosity caused by Wittgenstein's excessively inflated theory of syntax. After all, it takes some time to understand why Tractarian sentences are nonsensical. However, it would be good to have a theory explaining such a cardinal problem in an exegesis of a twentieth-century masterpiece.[10] Furthermore, Carnap had a tendency to accuse most philosophical classics of talking nonsense. Some of these accusations were based not on violations of the principle of verification, but on violations of logical syntax. As such, they are still valid to some extent, and, as we shall

[10] In this paper I can only suggest pretence as the mechanism behind understanding Tractarian nonsense. I'll devote another paper to the details of this view. My account is broadly in agreement with Cora Diamond's remarks on understanding Tractarian nonsense through imagination (2000, 157–160), but I would like to reinterpret imagination in terms of pretence. In her paper, Diamond observes that a nonsense sentence expresses no proposition, yet "to understand a person who utters nonsense is to go as far as one can with the idea that there is [a proposition]" (Diamond 2000, 157). I share with Diamond the framework of the austere view and a resolute reading of Wittgenstein's early thought, and I agree with some of her specific remarks on the topic of understanding a person who talks nonsense. For instance, she rightly recognises that we are neither inside nor outside that person's thought. At the same time, I appreciate McManus's criticism of Diamond's remarks on imagination: Diamond's suggestion is no more than a proposal that needs an elaboration that it has not yet received (McManus 2014, 171). It's not clear how Diamond reconciles the claim that nonsensical sentences have no content with her specific claims about the status of ethical and philosophical theses. How can Tractarian sentences exert any influence on the reader if these sentences convey no propositions? As far as I can see, Diamond's proposal amounts to a correct suggestion of a proposition-like structure of the imagination (Diamond 2000, 157, the words quoted above), and the false idea that attributions of nonsense are themselves nonsensical (Diamond 2000, 157–158). (Of course, Diamond's paper offers some fundamental views on the ethical dimension of the *Tractatus* and a penetrating critique of nonsensical sentences that "have something, something true, but unsayable"). I think that neither Diamond's (2010) remarks on the transitional sense of philosophical remarks nor Conant's views on the clarification of thoughts (2001, 60) are much more helpful in understanding nonsense than her remarks on "imagination".

see in the next paragraph, other concerns have recently been raised about the validity of much philosophical discourse. It would be good to know the mechanism of understanding philosophical works that are allegedly nonsensical.

A third domain of nonsense has recently been indicated by Herman Cappelen (2012, 2013). We, philosophers, like to think that our own assertions and statements are meaningful, and nonsense is a problem of the dead. Some historical philosophers made irresponsible claims and were irrevocably unclear about what their thoughts and claims meant; however, this does not happen now, in the twenty-first century (!), because the main characteristic of analytic philosophy is its clarity. Cappelen's Verbal Virus Theory (2012, 49–60) challenges the view that contemporary analytic philosophy is free of nonsense. He argues that the term 'intuition' and its cognates are semantically defective because they fail to have content, thus leading to sentential nonsense, assertoric nonsense, and cognitive nonsense.[11] Hence, for example, this sentence (5) is semantically defective:

(5) Intuitively, the Twin Earth liquid is not water.

'Intuition' is one of the most widely used philosophical terms of the second half of the twentieth century, but how did a meaningless term come to infect so much work? Here comes the virus part of the theory: as stated by Cappelen (2012, 50), 'Philosophers' use of "intuition" is a kind of intellectual/verbal virus (or tick) that started spreading about thirty to forty years ago'. What is the source of this virus? There are some clues, but this issue requires more historical analysis (Cappelen 2012, 56–57, Hintikka 1999).

Cappelen doesn't limit himself to claiming that only the term 'intuition' has such a miserable fate. No, its fate is far more common in contemporary philosophy. Cappelen postulates that the terms 'semantics', 'pragmatics', 'a priori', 'a posteriori', 'justification', 'causation', 'evidence' and 'person' also fall in the category of semantic defectiveness (Cappelen 2012, 60, fn.

[11] I borrowed this classification from (L. J. Keller and J. A. Keller 2021). *X* is a sentential nonsense if and only if *x* is a sentence that lacks content. *Y* is assertoric nonsense if and only if *y* is an assertion that lacks content. *Z* is cognitive nonsense if and only if *z* is a thought (belief, hope, desire, etc.) that lacks content. See Cappelen (2013, 26) for his own division of defective types of content.

10).[12] So, in his view, nonsense is quite widespread in contemporary philosophy. There are several reasons why these concepts are flawed. Roughly speaking, each of these terms is a theoretical term that has been defectively introduced. A term *T* is defectively introduced if it has one of the following characteristics (Cappelen 2012, 52):

- *T* has no agreed-upon definition among practitioners of a discipline.
- There is no agreement among practitioners of a discipline regarding which cases constitute core paradigms of the extension of *T*.
- There is no agreed-upon theoretical role that *T* (or *T*'s extension) plays in a discipline.
- There is considerable disagreement and dispute within a discipline about *T* itself.

This is only meant to be a rough outline, and it was later reformulated (Cappelen 2013, 38–40). However, it gives the reader an idea of what can be problematic about these terms. It's not surprising that terms such as 'semantics' or 'a priori' are used differently by scholars, and the discussion about what they mean is ongoing. It's not obvious how such terms acquire meaning in philosophical discourse, because experts disagree about their content. One possibility is that they inherit meaning from their natural language counterparts (Cappelen 2012, 27). I think that Cappelen's view is plausible for those terms that don't have a clear usage in natural languages. Such terms are 'intuition', 'semantics', 'pragmatics', 'a priori' and 'a posteriori'. They don't have fixed use in ordinary language and they seem to satisfy Cappelen's diagnostics.[13]

Cappelen (2013, 36) doesn't think that philosophers are particularly prone to nonsense, and he describes the practice of philosophers criticising

[12] Van Inwagen (1980) claims that philosophers' use of the word "body" is nonsensical, since there is no correct definition of this term. Furthermore, he suggests that such a definition isn't possible, and philosophers should give up using this term to avoid talking nonsense (cf. Tye 1980).

[13] Cappelen (2012, 29-48) argues convincingly that 'intuition' and its cognates don't have a common and uniform meaning in English. They can refer to a variety of states, events and things, and their meaning is context-dependent.

other philosophers for talking nonsense as somewhat objectionable. Be that as it may, he thinks that nonsense is generally more common than we think, and there are some good reasons for this. Speakers are fallible in respect to grounding facts of content. It may turn out that a term we use was introduced in a defective way, or that it was defectively transmitted through a chain of communication.[14] Cases of defective introduction include a lack of appropriate naming, demonstration, or intention on the part of the person introducing a term. I think language usage abounds with such cases. One such case has recently been discussed in the literature on demonstrative thoughts and various illusions of thought (L. J. Keller and J. A. Keller 2021, O'Brien 2009, Stojnić and LePore 2020). On the other hand, even if defective transmissions aren't widespread, they can happen. If communication is effective, it is so because it is carried out carefully. This condition isn't always met, and semantic failures occur.

To sum up, there are three distinct areas of potential philosophical investigation: I) fictional discourse; II) historical examination of Wittgenstein's, Carnap's, and others' views; III) illusions of thought as well as more contemporary discussions of notions such as "intuition", "semantics", "pragmatics", etc. I have chosen to group these areas in this way because all these groups are heterogeneous and propositions belonging to them may be semantically defective for different reasons: group I) consists of, among other things, category mistakes, poems made up of neologisms, and sentences that describe impossible states of affairs; group II) isn't uniform either. According to some commentators, nonsensical Tractarian sentences are violations of logical syntax (Glock 2004, Hacker 2003). Other scholars deny this and claim that they are just mere nonsenses (Conant 2003, Diamond 1978, 1981). Thus, depending on our interpretation of Wittgenstein, we may see his relation to Carnap's position on nonsense, which focused on violations of logical syntax, differently. The works of Hegel, Descartes and other philosophers may contain nonsense of other kinds. Van Inwagen (1980, 285), for example, claims that Descartes' *Meditations on First*

[14] The distinction between failures of introduction and failures of transmission is presumably not disjoint. I can unsuccessfully name someone and then pass this name to someone who, in turn, will transmit it 'incorrectly' (e.g., because she mishears me). J. J. Keller and J. A. Keller (2021) consider some possible scenarios of failures.

Philosophy contains nonsensical passages. Van Inwagen's work may suggest that these nonsenses are of the Cappelenian type, i.e., they result from a lack of agreement among experts on the correct definition of a term, or from a failure to introduce a term correctly. III) covers cases of a different kind from the Carnapian type. They have nothing to do with logical syntax and are closely related to the introduction of theoretical concepts. Cappelen believes that philosophy may have an important, purely cognitive function, but he claims that some terms in philosophy have not been properly introduced and thus result in nonsense. Nevertheless, other disciplines may produce nonsense of the same kind. In contrast to Cappelen, both Wittgenstein and Carnap were convinced that philosophical discourse is inherently flawed and philosophical errors are of a radically different kind from those in other domains.

I believe that this section has shown that there are some contexts in which we seem to we accept various kinds of semantically defective sentences. We engage in the practice of asserting and rejecting various sorts of nonsensical sentences in fiction, philosophy, and everyday life, thus it is hard to deny that nonsenses appear in the business of reasoning, thinking and judging. A pretence theory of understanding nonsense will explain how it is possible to engage in practices involving the use of nonsensical sentences in the specified domains. In the next section, I will describe the view which solves the riddle of understanding nonsense; also, I will present the rules of the game of make-believe that are responsible for the mechanism of understanding nonsense.

3. A pretence theory of understanding nonsense

I begin the presentation of my account with a few observations. First, I agree with the standard assumption that nonsensical sentences express neither propositions nor truth-conditions. To have meaning is to express some content, so nonsensicality must amount to a lack of content. This assumption is sometimes rejected (Sorensen 2002). I also accept the view that nonsensical sentences do not stand in logical relations. Only propositions imply other propositions or are entailed by other propositions. This view is also sometimes rejected (McManus 2014). Furthermore, I deny that nonsensical

sentences have logical forms (or partially interpreted logical forms). Oza (2022) claims that logical forms characterise sentences in general, and he argues that this view is necessary to explain the mechanism of understanding nonsense.[15] In my opinion, uninterpreted sentences are just physical objects and don't have logical properties. Sentences acquire logical forms if and only if they have content: they express propositions or truth-conditions, *etc.* Contents have logical forms, and we can derivatively ascribe logical forms to sentences. Sometimes, two or more sentences express the same content. In such cases, different propositions have the same logical form. The account that only propositions have logical forms seems to me no less standard than the first two assumptions.[16] For example, if one accepts Donnellan's (1966) view of the semantics of referential and attributive uses,

[15] There is an additional complication in Oza's (2022) views on the relation between pretence and logical form. According to him, we pretend that a nonsensical proposition expresses a logical form, but the pretended logical form is the one that the nonsensical sentence really has. This is not as unusual as it sounds: in fact, it is quite common for us to pretend to have some properties that we really have (see Langland-Hassan 2014, 11). For example, if I were to pretend that I am Wittgenstein, I would have to pretend that I am an Austrian, a philosopher, a human being, that I have a body and two legs, and that I have lived in Vienna for some time. But in fact I am a philosopher, a human being, I have a body and two legs, and I have lived for some time in Vienna. In contrast to Oza's position, my theory holds that we make believe that a nonsensical sentence expresses a logical form that is merely suggested by the nonsensical nonsense. Oza's stance helps him to show the difference between his view of nonsense and Diamond's position. He also believes that we can understand the logical form of a sentence that we don't understand (Oza 2022, 16–17). I disagree with this. I understand neither the meaning nor the logical form of the sentence "Postupně se vynořoval ucelený obraz genealogických souvislostí" because I don't speak Czech (I found it on Jaroslav Peregrin's website). Oza's view is well summarised in this fragment (Oza 2022, 20, fn. 29): "Given the role of form in my account, a string which is totally empty of form – say 'xg7*12d' – will, without further set-up, not be the subject of a pretence. Thus, 'it is nonsense that xg7*12d' can only be read metalinguistically. I think this is the correct result. To the extent that a nonsense sentence has some (even partial) syntactic form, there is the possibility of a pretence and a non-metalinguistic reading of the nonsense-attribution".

[16] I think it is shared by Higginbotham (1993) and LePore and Ludwig (2002), to whom Oza refers in other parts of his paper. In the context of logical structure of

then one must admit that the logical form of the sentence 'Smith's murderer is insane' depends on whether our intention was to express a truth-condition involving a quantifier expression or a truth-condition involving a singular term. Clearly the proposition remains the same, but its logical form is different. Moreover, if one accepts a wide range of contextual dependencies, then one has to acknowledge a greater degree of independence between a sentence and its logical form (Travis 1994, 2017). At present, contextualism seems to dominate in the dispute over the scope of contextual dependence.

Nonsensical sentences don't express propositions; they don't have a logical structure and they don't stand in logical relations. So, what is the mechanism behind understanding nonsensical sentences? It's pretence. When a speaker communicates using the nonsensical sentence s_1, she expresses the make-believe proposition p_1 within a game of make-believe. A hearer understands her correctly if and only if she grasps the make-believe proposition expressed by the speaker. More precisely, the hearer understands the nonsensical sentence s_1 if and only if she grasps the make-believe proposition p_1. There is a game of make-believe that specifies the proposition expressed by the nonsensical sentence.[17] Furthermore, when an agent derives the nonsensical sentence s_2 from the nonsensical sentence s_1, she pretends that the proposition p_1, which is expressed by the sentence s_1 in our game of make-believe, implies the proposition p_2, which is expressed by the sentence s_2 in our game of make-believe. The nonsensical sentence s_1 implies the sentence s_2 if and only if there is the pretended proposition p_2, which is entailed by the pretended proposition p_1 (in our game of make-believe).[18] A logical entailment is successfully communicated if and only if

sentence, he refers to the book by Armour-Garb and Woodbridge (2015), but these authors explicitly say that a hearer understands a logical form of some semantically defective sentences (e.g., a Liar sentence), but not full-blown nonsense such as 'colourless green ideas sleep furiously' (Armour-Garb and Woodbridge 2013, 849, fn. 6, 2015, 158, fn. 13). Thus, Armour-Garb and Woodbridge's view differs from Oza's position.

[17] In a sense that will be explained in a moment. For now, let's stick to a more informal way of speaking.

[18] There is a multitude of games of make-believe. Every inference and truth within pretence is relative to a game of make-believe. I think that works of fiction (books, plays, paintings) are games of make-believe, but for the purposes of this paper this

the hearer grasps the logical entailment make-believedly expressed by the speaker. Finally, a speaker pretends that the proposition p_1, which is make-believedly expressed by the nonsensical sentence s_1, is of the logical form LF_1; to grasp the make-believe proposition correctly, the hearer should correctly recognise the make-believe logical form.

Propositions in games of make-believe don't inherit logical structure from nonsensical sentences. Logical structure is another element that is pretended in a game of make-believe. The reason for this is well known: sentences (especially nonsensical ones) don't have a logical form. A make-believe proposition doesn't have to reflect the logical form of a sentence because nonsensical sentences can have incoherent logical forms that are not instantiated in the world. If the rules of logical syntax are supposed to specify prohibited combinations of logical categories (Hacker 2003, 7), then prohibited combinations should occur neither in propositions nor, hence, in language, which is a set of contents of sentences and speech acts. Category mistakes lead to nonsensical sentences that don't express propositions and have no logical structure. Logical forms which are forbidden by the rules of logical syntax do not occur as such. Let's take sentence (6) as an example:

(6) Chairman Mao is rare.

It is claimed that (6) is nonsensical (Dummett 1983, 51). 'Chairman Mao' is a name-like expression, and it requires a first-order predicate in order to be meaningfully combined (e.g., is bold, is a man). 'Rare' is a second-order predicate, and it can be ascribed to first-order predicates (e.g., an honest politician, gold).[19] (6) is supposed to combine two incompatible logical forms: 'a is φ' and 'ψ is X'.[20] (6) is a nonsensical sentence and doesn't express any proposition. In particular, it doesn't express the proposition of the logical form that combines 'a is φ' and 'ψ is X'. So, there is no proposition of this logical form. There are only some sentences which have a surface grammar in a form that combines 'a is φ' and 'ψ is X'. In my view,

assumption is unnecessary. There's no need to acknowledge more than the existence of children's games of make-believe and the like (Evans 1982, Walton 1990, 2015).

[19] See Conant's discussion of this example (2002, 395–398, 403–405).

[20] Lowercase Greek letters stand for first-order concepts, while uppercase Greek letters stand for second-order concepts.

the logical structures that are pretended in a game of make-believe are based on the surface grammar of sentences. The pretended proposition expressed by a nonsensical sentence usually has the logical form that it suggests. It's not the same logical structure because nonsensical sentences have no logical structure. Furthermore, it's a regularity that the surface grammar of a sentence is often the same as the logical form of the proposition, but there are exceptions. If someone were to express the nonsensical sentence 'The King of England is a prime number', then the proposition expressed in a game of make-believe would involve a quantifier-like statement.[21] The surface grammar of the sentence is different from the logical form of the expressed proposition.

When it comes to understanding nonsense, pretence functions as a stable mechanism that is exploited by a language user. The user simply exploits the appropriate convention of make-believe to achieve her communicative aims in accordance with the existing norms. She does not need to intend to pretend, nor does she need to be aware that she is pretending. The opposite appearance may stem from the ordinary meaning of the term "pretence". However, intuitions that link pretence with an intentional or deliberate act are incorrect (Armour-Garb and Woodbridge 2015, 63; Walton 1990, 38; Yablo 2001, 97). A pretence theory is concerned not with the mental states and introspection of speakers but with the linguistic functioning of some parts of discourse. By analogy, a theorist of fiction does not claim that an artist or her audience are aware of any pretence going on, and even less does a mathematician think that pretence is necessary to make any claims in his beloved field of inquiry. Walton even analyses dreams (1990, 43–50) and metaphors (1993) as types of pretence. So, what does the pretence episode look like from the speaker's point of view? Her knowledge is dispositional and practical. She proceeds, as Armour-Garb and Woodbridge (2015, 71) suggest, from the general assumption that names and designation expressions refer to objects and that predicates describe objects. She does not feel compelled to stop speaking in certain ways when confronted with putative obstacles to her practice, and she does not take a stance on how her sentences work. She simply uses these propositions as a tool to make

[21] Assuming the attributive reading of the description.

assertions. Furthermore, the speaker is not required to keep the rules of make-believe consciously in mind. In general, the rules of make-believe are not explicitly agreed or formulated, and the participants in a game of make-believe may not be aware of them. However, they are internalised by speakers: the rules of make-believe prompt the actions of agents, and in this sense they are operative in practice because speakers act upon them. For this reason, we have to observe the rules of make-believe in the practice of speaking a language in order to formulate them explicitly. Nevertheless, from the theoretical point of view, we should consider that the speakers are engaged in a game of make-believe and understand these fragments of discourse as involving pretence.[22]

There is a standard conception of games of make-believe, but the details of how to understand them vary (Armour-Garb 2015, Crimmins 1998, Evans 1982, Oza 2022, Walton 1990, 2015). The rules of make-believe can be divided into direct and indirect ones. Direct rules specify outright the basic principles of the game of make-believe; they include props and other stipulated pretences. Indirect rules prescribe what is make-believe based on the direct rules and real-world conditions, and they can be divided into principles of generation and recursive principles. Thus, the rules of a game of make-believe typically include:

a) Props,

b) Direct principles,

c) Indirect principles: Principles of generation (PG) and Recursive principles (RP).

a) and b) establish pretences that are expressly make-believe. For instance, in a children's game of make-believe they establish that a lump of mud is a biscuit and that some gestures towards the mouth are an act of eating the biscuit. Thanks to c), participants of a game of make-believe can draw upon their general knowledge of the world to generate the propositions, facts and events prescribed by the game. (PG) and (RP) aim to make the make-believe world as similar to the real world as the direct rules allow. For

[22] On this, see Armour-Garb and Woodbridge (2015, 63, 71–72) and Walton (1990, 38–40; 1993, 53).

example, in the absence of any indication to the contrary, we assume that all the protagonists of a novel eat, sleep, have blood in their veins, have ambitions, dreams, and so on. We assume that the fictional world is quite similar to our own. Of course, there are some differences between pretence episodes and their real-world counterparts.[23] In sum, games of make-believe set out local conventions that govern how some features of the actual world are to be mapped to some make-believe circumstances.

The proposed theory of understanding nonsense is semantic. The props for this game of make-believe are all linguistic expressions (unlike, for example, children's make-believe games, but similar to the rules of fiction). The rules are as follows[24]:

The Game of Make-Believe of Understanding Nonsense

A. Props:

The props for the game are the following types of linguistic expressions: names, designation expressions (including demonstratives and definite descriptions), predicative expressions (ascribing properties, relations *etc.*) and other types of linguistic expressions (quantifiers, logical connectives, *etc.*).

The following pretences are prescribed for these types of expressions.

(A.1) Every name or designation expression has a bearer.

(A.2) Every predicative term expresses a property or a relation.

(A.3) Every other expression has a semantic value.

[23] See Gendler (2003) on this topic. The differences have an impact on the way viewers receive fiction, especially the way they are emotionally engaged in the cinema (Gendler 2000, Liao and Gendler 2020, supplement: *Puzzles and Paradoxes of Imagination and the Arts*).

[24] Oza (2022, 18) proposed a set of pretences for understanding nonsense. However, the rules formulated in his work are not general enough to generate pretences for different cases of understanding nonsensical sentences. Oza's rules provide no more than an illustrative description of the pretences of the sentence 'Goodness is hexagonal'.

B. Direct principles:

(B.1) It is to be pretended that predicative expressions describe objects (denoted by names) as having or lacking some properties.

(B.2) It is to be pretended that sentences formed from props express propositions.

(B.3) It is to be pretended that sentences formed from props have logical form.

(B.4) It is to be pretended that sentences formed from props stand in logical relations.

(B.5) The pretences displayed in sentences formed from props are prescribed if and only if they have been asserted or used in other speech acts.

C. Indirect principles:

The principle of generation:

If P is true, and if there is no set of make-believedly true sentences $Q_1 \ldots Q_n$ such that if $Q_1 \ldots Q_n$ were true then P would not be true, then P is make-believedly true.

Recursive Principle:

If $P_1 \ldots P_n$ is a set of make-believe truths, and the counterfactual 'If $P_1 \ldots P_n$ were true, then R would be true' is true, and there is no set of make-believe truths $Q_1 \ldots Qn$ such that the counterfactual '$Q_1 \ldots Qn$ were true, then R would not be true' is true, then R is make-believedly true.[25]

(B.5) sets the condition for engaging in a game of make-believe of understanding nonsense. It is quite liberal, but it indicates that, in general, not much is required of a language user in order to engage in the prescribed make-believe. It is enough to read sentences from a novel to engage in the pretence prescribed by it. If there is an appropriate set of conventions, then

[25] How to correctly formulate (PG) and (RP) is, to some extent, an open question. It may depend on the subject of a game of make-believe. See the discussion in Walton (1990, 144–161).

the mere fact of using props (e.g., words, sentences, etc.) is enough to engage in some discourse.[26] Furthermore, Evans (1982, 256) notes that (PG) and (RP) are not sufficient to introduce make-believe propositional attitudes. It is generally not true that if one believes the truth of 'The pot is drinking coffee at t_1', then one believes the truth of 'the pot is not thirsty at t_2'. People tend to believe that pots don't drink coffee and pots aren't thirsty. Or, strictly speaking, people tend to believe that the sentence 'The pot is drinking coffee at *t*' is nonsensical and the sentence 'The pot is not thirsty' is nonsensical. Evans (1982, 257) proposes two principles for incorporating (some) propositional attitudes into a game of make-believe ('*()* stands for 'it is make-believedly the case that'):

(x) (If x believes that $*P*$, then $*x$ believes that $P*$)

(x) (If x intends that $*P*$, then $*x$ intends that $P*$).

The general rules of the game of pretending to understand nonsense prescribe pretences for particular instances of nonsensical sentences. For example, the rules prescribed for (1) are:

Props:

The props are the terms 'the pot', 'is', 'drink', and 'coffee'.

Direct principles:

The sentence 'The pot is drinking coffee at t_1' expresses the proposition that the pot is drinking coffee at t_1. The make-believe proposition has the logical form of '*a* is φ-ing at t_1'.

Indirect principles:

The truth 'The pot is drinking coffee at t_1' entails the truth 'the pot is not thirsty at t_2'.

[26] See Evans' (1982, Ch. 10) discussion of perceptual illusions. See also Armour-Garb and Woodbridge's (2013, 846, fn. 1, 2005, 67) rather general formulations of (what I take to be) the condition for engaging in some kind of pretense. I have benefited from their discussion of the rules of games of make-believe in these works.

Some utterances of nonsensical sentences seem to be semantically defective from the very beginning (e.g., "Julius Caesar is a prime number"), while others don't (e.g., "A picture is a fact"). What they have in common is that, taken literally, they do not express a proposition[27]. The main reason why a nonsensical sentence is defective is the semantic fact that it expresses no content. However, the fact that a sentence expresses no content may be unknown to the speaker who has decided to use it. On the other hand, if we follow the literal reading of nonsensical sentences, then any use of them would turn out to be semantically infelicitous. An appeal to semantic redirection involving pretence helps us to avoid ending up with semantically infelicitous assertions. Thanks to this appeal to pretence, we can explain how a semantically defective fragment of language serves any serious purpose at all, thus solving some philosophical puzzles by recognising pretence at work. In other words, the nonsensical sentence s_1 does not express a proposition, so to use this sentence to assert something we have to find another way that would be governed by rules of make-believe and real-world conditions. What makes an act of pretence appropriate as a move in the game of make-believe are direct principles, indirect principles, and some assertions made by the participants. Furthermore, pretence is intrinsic to understanding nonsense in the sense that the typical linguistic functioning of nonsensical sentences involves an appeal to it because pretence-free uses lead to assertoric nonsenses.[28]

The rules of make-believe and real-world conditions generate some acceptability conditions associated with an utterance of the nonsensical sentence s_1.[29] These acceptability conditions make an utterance of s_1

[27] I use the word "literally" only for lack of a better word. Some of the connotations it can carry are problematic. For example, literal should not be taken here in direct opposition to metaphorical discourse. Armour-Garb and Woodbridge (2015, p. 6, fn. 21) use the phrase "at face value" instead of "literally". From my perspective, "at face value" can be even more problematic because surface readings of nonsensical sentences are diverse.

[28] In this section, I have made use of some of the terminology introduced by Armour-Garb and Woodbridge (2015, chapters 1 and 2).

[29] In his otherwise critical article on the theory of pretence, Mark Richard (2013, 191) notes that the pretence account is not an account of what is said by the uses

appropriate or inappropriate; however, this sentence is still semantically defective (nonsensical) because it does not express a proposition. In practical terms, this means that if we use nonsensical propositions in our premises, then we must (in the end, in a serious mode) discard any conclusions based on the assumption that they are true. On the other hand, we can talk about an understanding of nonsensical sentences because the rejection of nonsensical sentences and of reasoning with nonsensical sentences is parasitic on pretence: it exploits the props, the direct principles, the indirect principles and the real-world conditions. When we reject nonsensical propositions, we exploit not only the fact that within a game of pretence there is a clear difference between being able to follow inferential relations and not being able to follow them, but also the fact that within a game of pretence there is a clear difference between the appropriateness of some propositions and the inappropriateness of other propositions.

In the next two sections, I will describe rival theories of understanding nonsense and show why they are unsatisfactory. Section 4 discusses Oza's (2022) pretence account of nonsense and explains why my pretence theory of understanding nonsense is more promising. Section 5 examines how L. J. Keller and J. A. Keller (2021) interpret the language of thought to make room for understanding nonsense.

4. Logical form and the characterisation of nonsense

I would like to begin a discussion of existing views from Oza's (2022) paper on theories of nonsense. Oza correctly recognises the role that

of some fragments of discourse. It does not assign to such uses anything like a Russellian proposition, a Fregean thought, or a set of possible worlds. Rather, it is a proposal about the truth conditions of utterances which purports to assign to each possible utterance of a fragment of discourse (paired with a pretence) a necessary and sufficient condition for the utterance to be correct. I don't like the use of the word "truth-conditions" since only meaningful sentences have truth-conditions, but Richard's suggestion is correct. According to this suggestion, pretence isn't just a substitution of one proposition for another. Language involves a variety of uses, and there are a variety of satisfaction conditions for those uses. Pretence involves appealing to one type of such use and its conditions of acceptance.

pretence can play in understanding nonsense. He appreciates the link between Diamond's discussion of imagination and pretence and notes the importance of these problems for the interpretation of Wittgenstein's Tractatus. As usual, Oza claims that pretence need not involve an intention to pretend. I think, however, that his position is ultimately untenable.

Oza (2022) maintains that a theory of nonsense must satisfy two requirements. The first one is the engagement constraint, which says that a theory of nonsense should explain how one can use nonsense in certain kinds of reasoning. Oza claims that the austere conception of nonsense fails to satisfy the engagement constraint, since it can only read nonsense ascriptions in a metalinguistic way.[30] The metalinguistic reading says that (7) must be read as (8).

(7) It is nonsense to say that a picture is a fact.

(8) The sentence 'a picture is a fact' does not express a thought.

The second requirement is the austerity constraint, which states that nonsensical sentences must not express propositions (or truth-conditions, or thoughts, *etc*). This constraint is inconsistent with the so-called no nonsense view (Bradley 1978, Prior 1954, Magidor 2009, 2013). According to this account of nonsense, sentences describing category mistakes such as (1) are not nonsensical but are necessarily false. Consequently, (1) expresses the proposition that the pot is drinking coffee and describes the state of affairs that the pot is drinking coffee. However, this proposition is false in every possible world, and there can be no state of affairs that it describes. Oza claims that this view of nonsense implies that the concept of nonsense loses its critical force. If nonsensical sentences express propositions, then they can play the same role as sense. Thus, there is no conceptual difference between nonsense and sense.

[30] The austere conception of nonsense holds that there is only one kind of nonsense, and that this kind of nonsense has its source in the lack of meaning in one of the components of a sentence. In contrast to the standard view of nonsense, it says that the nonsensicality of (2) is not the result of a violation of the rules of logical syntax. For a discussion, see (Conant 2001, 2002, Dain 2006, 2008, Diamond 1978, 1981, 2005). For a recent defence, see my (Bogucki 2023).

Oza claims that the only view which satisfies both the austerity constraint and the engagement constraint is the pretence account of nonsense which he proposes. His view has it that nonsensical sentences do not express propositions, but that a typical user of language who engages with nonsense pretends that a nonsensical sentence express a proposition. This view satisfies both the engagement constraint and the austerity constraint. A pretence account of nonsense satisfies the first constraint, since an agent operates with a make-believe proposition in reasoning. The second constraint is satisfied because Oza accepts the view that category mistakes are nonsensical. The two constraints are clearly satisfied, so what's wrong with Oza's framework?

The main problem with Oza's position is that it mischaracterises the scope of the theory of understanding nonsense. The austere conception of nonsense, the standard view of nonsense, and the no nonsense view are substantive theories that describe the properties of a class of sentences. These theories debate the status of category mistakes, what semantic properties these sentences have, and what the source of these semantic properties is. The austere conception and the standard view agree that category mistakes are nonsensical, but they differ on how to explain this fact. The no nonsense view claims that category mistakes are meaningful and tries to explain their special semantic status in a different way. A theory of understanding nonsense is a higher-level theory, and it can be reconciled with any theory of nonsense.[31] The theory of understanding nonsense aims to explain

[31] Any theory that recognises some sentences as nonsensical. The no nonsense view says that category errors are meaningful, but it does not exclude other kinds of nonsense. Jabberwocky cases (nonsense made up of neologisms), terms that have no definition or clear theoretical role (cases discussed by Cappelen), and illusions of thought in general (discussed by Cappelen (2013) and J. A. Keller and L. J. Keller (2021)) are compatible with this view of nonsense. In fact, Cappelen (2013. 86, fn. 5) is quite sympathetic to the no nonsense approach, and he doesn't see any tension with his other examples of nonsensical sentences. Furthermore, there are simple cases of nonsense (e.g., "Mark is xywwww", "The cat on the mat is phlump") that contain a word without meaning. These sentences can be used because they can be the subject of nonsense attributions, e.g., "David believes that the cat on the mat is phlump". More generally, simple cases of nonsense can also be used as assertions and lead to assertoric nonsense.

how we understand nonsensical sentences, how we can reason with nonsensical sentences, and how to explain the difference between understanding nonsense and understanding sense. As we have seen above, theories of nonsense discuss a different class of problems. One can be a proponent of the austere conception and accept the pretence theory of understanding nonsense. One can embrace the standard view of nonsense and reject the pretence theory of understanding nonsense at the same time. Furthermore, if one accepts the austere view and the pretence theory of understanding nonsense, then both of Oza's constraints are satisfied. The explanation for this fact is simple: according to the austere view, nonsensical sentences do not express propositions, hence the austerity constraint is satisfied; on the other hand, the pretence theory of understanding nonsense provides an explanation of how engagement with nonsense is possible. Oza's mischaracterisation of theories of understanding nonsense has implications for his whole framework. In fact, he believes that his view is a 'pretence account of nonsense' rather than a theory of understanding nonsense. Next, I will show that Oza's position also mischaracterises the concept of nonsense itself.

Oza thinks that the pretence account of nonsense is on the same level as the austere view of nonsense, hence he wants to show that the austere view doesn't satisfy the engagement constraint. His main worry is that, according to the austere conception, 'when we produce nonsense we aren't using our conceptual capacities. But our engagement with nonsense draws precisely on these capacities' (Oza 2022, 7). This has some important consequences for our understanding of language. Firstly, there is no explanation of the difference between sentences (7) and (9).

(7) It is nonsense to say that a picture is fact.

(9) It is nonsense to say that das Bild eine Tatsache ist.

A proponent of the austere view can't say that the difference is simply knowledge of a language since she must accept the metalinguistic reading of (7) and (9). The metalinguistic reading rejects the appeal to concepts. Secondly, there is no explanation of how a language user understands a sentence that she has never heard before (Oza 2022, 9). The main motivation for the austere view is a reading of Frege's context principle, which says that the meaning of a word is its contribution to the meaning of the

particular sentence in which it occurs. This reading implies that a word makes no general contribution to a sentence.

First of all, I think the difference between (7) and (9) is understandable if we combine the austere view with the pretence theory of understanding nonsense, and the explanation of this difference is the same as Oza has given. However, Oza's concern that the austere view is incapable of accounting for our conceptual capacities has some deeper motivations that are also present in the worry about the productivity of our language. These concerns are based on the view that the austere conception is incompatible with the existence of general rules of language.[32] I think these worries are based on a misunderstanding of the austere view of nonsense. Historically speaking, Diamond has acknowledged that there are semantic and syntactic rules that tell us how to understand the meanings of words and sentences (Diamond 1978, 199–201, 1981, 19–20). These general rules underlie our understanding of sentences. For instance, if I know the rules of English, then I know that the meaning of the sentence 'Venus is more massive than Mercury' is determined by the meanings of 'Venus', 'Mercury' and the relational expression 'x is more massive than y' (Diamond 1981, 19). Furthermore, the knowledge of these general rules tells us that 'Venus is more massive than Mercury' has some elements common with 'Venus is less massive than Saturn', but not with 'Die Venus ist weniger massiv als der Saturn'. I think that the existence of general rules of language is sufficient to explain the difference between (7) and (9) without resort to a pretence theory of understanding nonsense. (7) is more informative than (9) because it makes use of and requires our knowledge of the general rules of different languages.

Another difference between Diamond (the proponent of the austere view) and Oza (the advocate of the standard view of nonsense) is that Diamond (1981, 19–20) claims that understanding the rules of language is conditional. The conditional understanding of rules means that we can use the general rules of our language to characterise the meaning and structure

[32] Glock (2004, 2015) and Liptow (2018) have also raised this objection. I cannot discuss the details here, but I have elaborated on this problem and responded to it in (Bogucki 2023, 19–21). The objection has something to do with the Glock's and Liptow's reading of the principle of compositionality.

of a sentence, but only conditionally. If a sentence is nonsensical, then there is no way of identifying its syntactic and semantic elements.[33] I have argued elsewhere that this aspect of the austere view is not incompatible with the principle of compositionality and productivity of language, hence I will limit myself to one observation.[34] The standard view of nonsense is also committed in some cases to the conditional understanding of rules of language, and the defenders of this view see no problem in appealing to it in the context of the principle of compositionality.[35] So, the conditional understanding of rules shouldn't be seen as a problem *per se*. The existence of semantic and syntactic rules underlies our knowledge of language, but sometimes the knowledge of rules tells us that a certain sequence of signs hasn't been used and should remain without semantic interpretation. Lack of interpretation is not a sign of absence of conceptual abilities.

I have already touched upon another problem with Oza's framework, namely the problem with ascription of logical form to sentences. In my view, only interpreted signs have logical structure, since meaningful sentences have logical form by virtue of their contents. Nonsensical sentences have no content since they don't express propositions (or truth-conditions, etc.); therefore, the only principled way of ascribing logical structure to nonsensical sentences is in terms of their surface grammar. The opposite view is a necessary ingredient of the standard view of nonsense, and thus of Oza's position. Proponents of the standard view ascribe logical form to nonsensical sentences and their constituents, and they maintain that a sentence is nonsensical because of the incompatibility of the logical categories of its

[33] For example, if we take the meaningful sentence "Venus is more massive than Mercury", we can determine that, according to the general rules of language, "Venus" refers to the planet Venus, "Mercury" refers to another planet Mercury, and "is more massive" expresses a relation between two objects. However, if the sentence in question is nonsensical, it is not possible to identify the logical elements because the term "Venus" can be used as a predicate in one sentence and as a proper name in another (see Diamond 1979, 209–211, 1981, 7–13, 19–21). The conditional understanding of semantic and syntactic rules fits well with the claim that nonsensical sentences don't have a logical form.

[34] See again my Bogucki (2023, 17–23).

[35] Glock (2015, 127–128). For a discussion, see Magidor (2009, 556–565).

constituents. (2) is nonsensical because the logical category of 'Julius Caesar' and the logical category of 'a prime number' are incompatible. Predicating an abstract property of a concrete object violates the rules of logical syntax. Nonsensical sentences have logical structures; thus, their constituents possess logical categories because the logical properties of sentences are the source of the exclusion of nonsensical sentences from language. As I said in the previous section, an attribution of logical form is problematic because sentences can express different propositions, and these propositions can have incompatible logical forms.[36] But there is another problem with this view that is again common to Oza's framework and to any standard view of nonsense: it is the problem of making sense that is too similar to nonsense. The challenge is that an explanation of why a sentence is nonsensical leads to attributions of properties to sentences that can only be made about meaningful thoughts and propositions.[37] Oza (2022, 10) claims that the no nonsense view cannot deal with this problem and notes that, according to this view, 'when a subject attempts to understand a nonsense sentence, they exercise the same capacities they would exercise in understanding the sentence's words and structure elsewhere'. However, Oza's position seems to imply the same difficulties. Typically, meaningful sentences express propositions because they consist of concepts and logical forms. According to Oza, nonsensical sentences also consist of concepts and logical forms. His view implies that meaningful and nonsensical sentences share common concepts and logical properties.[38] This similarity can lead one to the conclusion that

[36] I also observed in the previous section that we ascribe logical forms to nonsensical sentences which, in fact, do not exist in our language and are said to be incompatible with logical syntax. If these logical forms are incorrect, then they must also be excluded from sentences.

[37] Oza (2022, 10) formulates this difficulty as follows: '(…) there is some difficulty in explaining why a "thought" is meaningless without presupposing that it is meaningful'.

[38] The no nonsense view and Sorensen's account (2002) blatantly violate the austerity constraint because they claim that nonsense has contents. The standard view of nonsense denies that it expresses propositions, but it does not explain why this is so. It seems as if nonsensical sentences express some illogical thoughts. Hence, there is a reason to think that the standard view non-blatantly violates the austerity constraint.

'It's unclear, then, why a subject attempting to understand nonsense fails to grasp a thought: the subject's activity looks the same as in the good case' (Oza 2022, 10). James Conant (2001) has argued that such a problem is common to any rendering of the standard view of nonsense, and Oza's framework inherits this difficulty.

5. The language of thought hypothesis and nonsense

L. J. Keller and J. A. Keller (2021) propose another solution to the riddle of understanding nonsense. In the Introduction, I formulated the riddle in terms of propositions, truth-conditions, and other possible contents of sentences. So, if a sentence doesn't express a proposition, then it is hard to explain how an agent can understand this sentence. Typically, understanding a sentence consists in understanding the proposition that it expresses. L. J. Keller and J. A. Keller (2021) agree that there are some cases in which an agent believes nonsensical sentences. Furthermore, they argue that such an agent is usually justified in thinking that a nonsensical sentence contains some content. L. J. Keller and J. A. Keller's explanation of the phenomenon of understanding nonsense (and illusions of thought) is to reject the view that to understand a sentence is to understand the proposition that it expresses. Instead, they accept the language of thought hypothesis.

The LOT hypothesis is said to be a language-like system of internal representation.[39] According to it, beliefs and other propositional attitudes are not relations to propositions. All propositional attitudes are mediated. When an agent understands a sentence, she entertains a mental representation of the proposition expressed by this sentence. Propositions and other possible contents of sentences are mediated by our supposed internal language; they don't appear directly in a language user's mind since only their mentalese translations do so.[40] On the other hand, the contents of sentences

[39] Fodor (1975) and Schiffer (1981) are the *locus classicus* of LOT.

[40] The mind is divided into a number of boxes because we can have different propositional attitudes. To have a propositional attitude is to have a mental translation of the content of a sentence that describes our attitude (Schiffer 1981).

translated into mentalese determine the contents of our mental actions. L. J. Keller and J. A. Keller (2021) appeal to this feature of LOT in their solution to the riddle of understanding nonsense. They distinguish between 'beliefs', 'thoughts', *etc.*, understood as some acts of believing, thinking, and 'beliefs', 'thoughts', understood as some contents of these acts.[41] When I think the sentence 'Joe Biden is the current president of the United States', then there is the proposition that describes the content of my belief, and there is the act of thinking this content. So, what is going on when we are thinking nonsensical sentences? L. J. Keller and J. A. Keller (2021, 242) say:

> With this distinction [between the acts of thinking and their contents] in hand, we can say that while illusions of thought do not, by definition, involve thought contents, they may still involve beliefs$_s$ (and hopes$_s$ and desires$_s$). It's just that those beliefs$_s$, hopes$_s$, and desires$_s$—those thoughts$_s$—are empty. We can thus give a unified account of thinking, speaking, and writing nonsense: thinking nonsense involves actually thinking empty thoughts; speaking nonsense involves actually speaking empty words, and writing nonsense involves actually writing sentences that don't (actually) express contents.

L. J. Keller and J. A. Keller (2021) maintain that mentalese representations of nonsensical sentences have some contents, but these contents are empty. Every public language sentence has its mentalese counterpart, and the proposition expressed by the public language sentence determines the content of its mentalese counterpart. Nonsensical sentences are devoid of content, hence when they are represented in mentalese sentences, they don't give them any specific content. On the other hand, the act of thinking about a content is different from the content itself, so there is an act of thinking about a nonsensical sentence even though that sentence has no content. Does this solve the puzzle of understanding nonsense?

First of all, if someone doesn't accept the language of thought hypothesis for some other reason, the price of solving the puzzle of understanding is

[41] L. J. Keller and J. A. Keller claim that the act of believing or thinking is a concrete mental state or event. They mark these states with the subscript *s*.

quite high. LOT is a controversial view that proposes a substantive theory of mind and a non-classical account of propositional attitudes. It carries with it a number of ontological commitments. One may raise the concern whether it is worth appealing to LOT in order to solve a particular problem in the theory of language. Moreover, even if LOT is theoretically indispensable, it needs empirical confirmation, and the tests may turn out to be unsuccessful. Of course, LOT also has some important advantages, but it seems to me that the pretence theory of understanding nonsense has fewer and less serious commitments. For this reason, it is preferable to LOT, even if both views are equally successful in explaining the problem of understanding nonsense.

So, is LOT able to explain the riddle of understanding nonsense? In my opinion, the answer to this question is negative. Firstly, it is hard to understand what translations of nonsensical sentences into mentalese are. L. J. Keller and J. A. Keller (2021) claim that if a sentence is nonsensical, then its mentalese translation is also without meaning. (3) is nonsensical, so one's belief box contains its mentalese counterpart 'THE THEORY OF EVOLUTION IS EATING BREAKFAST'.[42] The explanans, namely the fact that one's belief box contains a nonsensical mentalese sentence, doesn't seem any more intelligible and self-explanatory than the explanandum – the fact that one believes a nonsensical sentence. What are mentalese translations of nonsensical sentences? L. J. Keller and J. A. Keller (2021) maintain that mentalese translations of nonsenses have contents that are empty. This explanation doesn't tell us much. Do mentalese translations also have empty logical contents? The notion of empty contents seems no more informative than the well-known concept of sentences without contents. It is hard to see any theoretical gain in postulating nonsensical mentalese sentences. Therefore, it seems to me that L. J. Keller and J. A. Keller (2021) should have spent more time elaborating on the utility of this concept.

Secondly, one might doubt whether there really are such things as mentalese translations of nonsenses. According to LOT, to believe that 'snow is white' is to have a mentalese translation of the content of the sentence 'snow is white' in one's belief box. So, if I believe that p, I have a mental

[42] It's L. J. Keller and J. A. Keller's (2021) convention to mark mentalese counterparts with capital letters.

representation of a proposition that *p*. The problem is that nonsensical sentences don't express propositions (or truth-conditions). Hence, there is no content to be translated into mentalese at all. What do we translate into mentalese in such a case? In the previous paragraph, I assumed that we could ignore this problem and postulate that we translate sentences, but sentences don't have semantics. They have to be interpreted, and they can only get content thanks to propositions or truth-conditions. LOT's mentalese sentences are supposed to have semantics, but they can't get them from semantically uninterpreted sentences. I think that some proponents of LOT are aware of this problem (Recanati 1997, Sperber 1997). Dan Sperber (1997) claims that one's belief box can only contain a meta-representation of the proposition *p* if one doesn't fully understand *p*. So, if a schoolboy doesn't understand the proposition that Cicero's prose is full of 'synecdoches' (because he doesn't understand the meaning of the term 'synecdoche'), his belief box can contain a meta-representation of this proposition but not its representation.[43] Recanati (1997, 91) agrees that '(…) a sentence cannot make its way into the mind (whether into the belief box or elsewhere) if it contains uninterpreted symbols'. A representation must be interpreted in some or other way, otherwise it is difficult to understand what it means to mentally entertain uninterpreted symbols. Half-understood sentences ('Cicero's prose is full of "synecdoches"') and nonsensical sentences are uninterpreted.

Recanati appeals to Kaplan's character-content distinction to solve the problem of understanding uninterpreted symbols[44]. Sometimes a representation of a proposition is determined by the character of an expression. Character is informative independently of content since it shows the route to the proposition, even if that proposition doesn't exist. Recanati postulates introducing a deferential operator which provides character to uninterpreted sentences. The deferential operator $R_x()$ applies to the symbol σ and yields a complex expression $R_x(\sigma)$. The character of $R_x(\sigma)$ takes us from a context in which the speaker tacitly refers to a particular cognitive agent

[43] The schoolboy believes that *the teacher says that Cicero's prose is full of 'synecdoches'*, but he can't entertain the proposition that *Cicero's prose is full of 'synecdoches'*.

[44] See Kaplan (1989) on the distinction between character and content.

x to the content which σ has for x. This solves the problem, since one's belief box doesn't have to contain the representation of the uninterpreted sentence or its meta-representation. This belief box contains a deferential representation that is the proposition-like character of the previously uninterpreted sentence. For instance, the character of the sentence 'Cicero's prose is full of "synecdoches"' takes us from the cognitive agent that is tacitly referred to (i.e., the teacher) to the content that the symbol 'synecdoche' – and thus the whole sentence – has for the teacher. Loosely speaking, the character says that the content of 'synecdoche' is whatever concept the teacher was referring to, and that concept – together with the rest of the constituents – determines the sentence.

In my opinion, Recanati's amendment is only plausible for sentences which have some natural candidates for a particular character of an expression. Typical cases of nonsense, like 'Julius Caesar is a prime number', don't easily submit to the character/content distinction. Another of Recanati's examples is the sentence 'The unconscious is structured like language', the character of which again points us to the author's intended concept. The difference is that, according to Recanati, the author of this *bon mot*, Jacques Lacan, didn't have a specific meaning in mind when he made this statement. Hence, the sentence has a character, but it doesn't have any content. The character of the sentence allows us to get the deferential representation. It is not clear what the character of the sentences 'The pot is drinking coffee' or 'Julius Caesar is a prime number' is supposed to be. In both cases, the utterer of the sentence wanted to express a proposition that doesn't exist. Character can't direct us to the proposition because the utterer knew that the proposition is semantically defective.[45] The riddle of nonsense is to find the proposition that is missing. Recanati's theory doesn't say how to find the lacking proposition, nor does it provide a plausible character for the most important cases of nonsense. His two examples are similar in that they emphasise the author's intention, but there is no such authorial intention in the most interesting cases. Furthermore, the community of users does not determine the expressed proposition. There is no publicly available

[45] Strictly speaking, the sentence is semantically defective and fails to express a proposition.

meaning.[46] The character of nonsense can't refer to a non-individualistically interpreted user. Thus, Recanati's solution may seem promising for half-understood sentences, but it fails in the case of full-fledged nonsense. Full-fledged nonsense doesn't have plausible characters, and it's even hard to conceptualise some loose descriptions of the candidates.

Finally, L. J. Keller and J. A. Keller's (2021) account cannot explain the role of nonsensical sentences in reasoning. Mentalese translations of nonsensical sentences inherit semantic properties from their contents. On the one hand, nonsensical sentences have no content, and mental translations can't get semantic properties from them. On the other hand, nonsensical sentences don't have logical relations and logical structures. Their mentalese translations should have the same properties, which is why it is hard to understand how we can use nonsensical mentalese sentences in reasoning and draw conclusions from them. Does the mentalese sentence 'the pot is drinking coffee at t_1' entail the mentalese sentence 'the pot is not thirsty at t_2'? As we have seen, in some circumstances nonsensical sentences appear in inferential connections and justify our actions. Moreover, L. J. Keller and J. A. Keller discuss sentences like 'Witches cast spells' which are also semantically defective.[47] People think that if witches cast spells, then witches are dangerous, and witches should be avoided since they can cause numerous misfortunes. The properties of empty mentalese sentences don't explain why these sentences are logically related. The explanation provided by nonsensical mentalese sentences does not seem to be any more helpful in such cases than their linguistic counterparts. The distinction between the act and the content of thinking doesn't help either. The act of thinking nonsensical sentences has empty content that can play no role in reasoning. The fact that mental realisations of this content are distinct doesn't change much.

[46] Contrast this with the half-understood sentence 'I have "arthritis" in my thigh', discussed by Recanati. In this case, a speaker doesn't know the public meaning of this sentence, but the language community collectively determines the content (Burge 1979).

[47] See L. J. Keller and J. A. Keller (2021, 227, fn. 6) for an explanation of the semantic defectiveness of this sentence.

6. Conclusion

In this paper I have explored the riddle of understanding nonsense. This riddle concerns the possibility of entertaining nonsensical sentences, judging their content, and drawing conclusions from them, given the fact that nonsensical sentences don't express propositions that are standard objects of propositional attitudes. In Section 2, I elaborated on the importance of this riddle. In Section 3, I proposed a pretense theory of understanding nonsense. In Section 4, I examined Oza's alternative framework, which is based on the standard view of nonsense and attributes logical forms to sentences. In Section 5, I discussed L. J. Keller and J. A. Keller's proposal to explain the riddle of understanding nonsense by means of the language of thought hypothesis. I think my considerations show that Oza's and L. J. Keller and J. A. Keller's proposals have some shortcomings that make my own view more favourable than their positions.

Disclosure statement

No potential conflict of interest was reported by the author.

Acknowledgements

I would like to thank Tadeusz Ciecierski, two anonymous referees for this journal and the audiences at the MLAG Seminar at the University of Porto, the Seminar Series in Analytic Philosophy at the University of Lisbon and the 11th Polish Philosophical Congress at the University of Łódź for their helpful feedback.

Funding

This work was supported by the National Science Centre, Poland under Grant 2022/44/C/HS1/00054.

References

Armour-Garb, Bradley., and James A. Woodbridge. 2013. "Semantic Defectiveness and the Liar". *Philosophical Studies* 164: 845–863. https://doi.org/10.1007/s11098-012-9915-6

Armour-Garb, Bradley., and James A. Woodbridge. 2015. "Pretense and Pathology: Philosophical Fictionalism and its Applications". Cambridge University Press.

Bogucki, Krystian. 2023. "A Defence of the Austere View of Nonsense". *Synthese* 201 (5): 1–30. https://doi.org/10.1007/s11229-023-04138-x

Bradley, Michael. 1978. "On the Alleged Need for Nonsense". *Australasian Journal of Philosophy* 56 (3), 203–218.

Burge, Tyler. 1979. "Individualism and the Mental". *Midwest Studies in Philosophy* 4 (1): 73–122. https://doi.org/10.1111/j.1475-4975.1979.tb00374.x

Camp, Elisabeth. 2004. "The Generality Constraint and Categorial Restrictions". *Philosophical Quarterly* 54 (215): 209–31. https://doi.org/10.1111/j.0031-8094.2004.00348.x

Cappelen, Herman. 2012. *Philosophy Without Intuitions*. Oxford: Oxford University Press UK.

Cappelen, Herman. 2013. "Nonsense and Illusions of Thought". *Philosophical Perspectives* 27: 22–50. https://doi.org/10.1111/phpe.12031

Carnap, Rudolf. 1959. "The Elimination of Metaphysics Through Logical Analysis of Language". In *Logical Positivism*, edited by Alfred J. Ayer, 60–81. New York: The Free Press.

Conant, James. 2001. "Two Conceptions of Die Uberwindung der Metaphysik". In *Wittgenstein in America*, edited by Timothy McCarthy, Sean C. Stidd, 13–61. Oxford: Oxford University Press.

Conant, James. 2002. "The Method of the Tractatus". In *From Frege to Wittgenstein: Perspectives on Early Analytic Philosophy*, edited by Erich H. Reck, 374–462. Oxford: Oxford University Press.

Crimmins, Mark. 1998. "Hesperus and Phosphorus: Sense, Pretense, and Reference". *Philosophical Review* 107 (1): 1–47. https://doi.org/10.2307/2998314

Dain, Edmund. 2006. "Contextualism and Nonsense in Wittgenstein's Tractatus". *South African Journal of Philosophy* 25 (2): 91–101. https://doi.org/10.4314/sajpem.v25i2.31436

Dain, Edmund. 2008. "Wittgenstein, Contextualism and Nonsense: A Reply to Hans-Johann Glock". *Journal of Philosophical Research* 33: 101–125. https://doi.org/10.5840/jpr_2008_8

Diamond, Cora. 1978. "Frege and Nonsense". In *Intention and Intentionality: Essays in Honour of G. E. M. Anscombe*, edited by Cora Diamond, and Jenny. Teichman, 195–220. Brighton: Harvester.

Diamond, Cora. 1981. "What nonsense might be". *Philosophy* 56 (215): 5–22. https://doi.org/10.1017/S0031819100049743

Diamond, Cora. 2000. "Ethics, Imagination and the Method of Wittgenstein's Tractatus". In *The New Wittgenstein*, edited by Alice Crary and Rupert Read, 149–73. Routledge.

Diamond, Cora. 2005. "Logical Syntax in Wittgenstein's Tractatus". *Philosophical Quarterly* 55 (218): 78–89. https://doi.org/10.1111/j.0031-8094.2005.00386.x

Diamond, Cora. 2010. "Inheriting from Frege: The Work of Reception as Wittgenstein Did It", In *The Cambridge Companion to Frege*, edited by Michael Potter, and Thomas Ricketts, 550–601. Cambridge: Cambridge University Press.

University Press: 550–601.

Donnellan, Keith. 1966. "Reference and Definite Descriptions". *Philosophical Review* 75 (3): 281-304. https://doi.org/10.2307/2183143

Dummett, Michael. 1983. *Frege: Philosophy of Language*. 2nd edition. London: Duckworth.

Evans, Gareth. 1982. *The Varieties of Reference*. Oxford University Press.

Fodor, Jerry. 1975. *The Language of Thought*. Crowell Press.

García-Carpintero, Manuel. 2019. "Semantics of Fictional Terms". *Teorema. International Journal of Philosophy* 38 (2): 73–100.

Gendler, Tamar. 2000. "The Puzzle of Imaginative Resistance", The Journal of Philosophy, 97(2): 55–81. https://doi.org/10.2307/2678446

Gendler, Tamar. 2003. "On the Relation between Pretence and Belief". In *Imagination, Philosophy, and the Arts*, edited by Mathew Kieran, and Dominic. M. Lopes, 124–141. New York: Routledge.

Glock, Hans-Johann. 2004. "All kinds of nonsense". In *Wittgenstein at Work*, edited by Erich Ammereller, and Eugen Fischer, 221–245. London: Routledge.

Glock, Hans-Johann. 2015. "Nonsense Made Intelligible". *Erkenntnis* 80: 111–136. https://doi.org/10.1007/s10670-014-9662-5

Hacker, Peter. 2003. "Wittgenstein, Carnap and the New American Wittgensteinians". *The Philosophical Quarterly* 53: 1–23. DOI: 10.1111/1467-9213.00292

Higginbotham, James. 1993. "Grammatical form and logical form". *Philosophical Perspectives* 7:173–196. https://doi.org/10.2307/2214121

Hintikka, Jaakko. 1999. "The Emperor's New Intuitions". *Journal of Philosophy* 96(3): 127–47. https://doi.org/10.5840/jphil199996331

Langland-Hassan, Peter. 2014. "What It Is to Pretend". *Pacific Philosophical Quarterly* 95 (1): 397–420. https://doi.org/10.1111/papq.12037

Lepore, Ernie., and Ludwig. Kirk. 2002. "What is Logical Form?". In *Logical Form and Language*, edited by Gerhard Preyer, and Georg Peter. Clarendon Press.

Lewis, David. 1978. "Truth in Fiction". *American Philosophical Quarterly* 15: 37–46.

Liao, Shen-yi., and Tamar. Gendler. 2020. "Imagination". The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), edited by Edward N. Zalta. Last

updated: January 22, 2019. https://plato.stanford.edu/archives/sum2020/entries/imagination/

Liptow, Jasper. 2018. "Wittgenstein's Argument for the Context Principle in the Tractatus". *Grazer Philosophische Studien* 95: 224–44. https://doi.org/10.1163/18756735-000037

Kaplan, David. 1989. "Demonstratives: An Essay on the Semantics, Logic, Metaphysics and Epistemology of Demonstratives and other Indexicals". In *Themes from Kaplan*, J. Almog, J. Perry, and Howard Wettstein, 481-563. Oxford University Press.

Keller, Lorraine. J., and John.A. Keller. 2021. "Saying Nothing and Thinking Nothing". In *Non-Being: New Essays on the Metaphysics of Nonexistence*, edited by Sara Bernstein, and Tyron Goldschmidt, Oxford University Press.

Magidor, Ofra. 2009. "Category Mistakes are Meaningful". *Linguistics and Philosophy* 32: 553–81. https://doi.org/10.1007/s10988-010-9067-0

Magidor, Ofra. 2013. *Category Mistakes*. Oxford University Press.

McManus, Denis. 2014. "Austerity, Psychology, and the Intelligibility of Nonsense". *Philosophical Topics*: 1–53.

Nolan, Daniel. 2021. "Impossible Fictions, part I: Lessons for fiction". *Philosophy Compass*: 1–12. https://doi.org/10.1111/phc3.12723

O'Brien, Lucy. 2009. "Mental Actions and the No-Content Problem". In *Mental Actions*, edited by Lucy O'Brien and Matthew Soteriou, 215-230. Oxford University Press.

Oza, Manish. 2022. "Nonsense: a User's Guide". *Inquiry*: 1–31. DOI: 10.1080/0020174X.2022.2107059

Prior, Arthur. 1954. "Entities". *Australasian Journal of Philosophy* 32: 159–68. https://doi.org/10.2307/2269122

Recanati, Francois. 1997. "Can We Believe What We Do Not Understand?". *Mind and Language* 12 (1): 84-100. https://doi.org/10.1111/1468-0017.00037

Richard, Mark. 2013. "Semantic Pretence". In *Context and the Attitudes. Meaning in Context, vol. 1*, 172-196. Oxford: Oxford University Press.

Russell, Bertrand. 1908. "Mathematical Logic as Based on the Theory of Types", *American Journal of Mathematics* 30: 222–62. DOI: 10.2307/2272708

Russell, Bertrand. 1910. *Principia Mathematica, vol. 1*. Cambridge University Press, 2nd edition published in 1997.

Ryle, Gilbert. 1949. *The Concept of Mind*. London.

Ryle, Gilbert. 2009. "Categories". In *Collected Papers, vol. 2*, 178–193. Taylor and Francis.

Schiffer, Stephen. (1981). Truth and the theory of content. In Herman Parret (ed.), *Meaning and Understanding*. Berlin.

Searle, John. 1975. "The Logical Status of Fictional Discourse". *New Literary History* 6: 319–332.

Sorensen, Roy. 2002. "Meaningless Beliefs and Mates's Problem". *American Philosophical Quarterly* 39 (2): 169–182.

Sperber, Dan. 1997. "Intuitive and Reflective Beliefs". *Mind and Language* 12 (1): 67–83. https://doi.org/10.1111/1468-0017.00036

Stojnić, Una., and Ernie Lepore. 2020. „Fodor and Demonstratives in LOT". *Theoria: An International Journal for Theory, History and Foundations of Science* 35 (1): 75–92.

Sullivan, Peter. 2003. "Ineffability and Nonsense". *Proceedings of the Aristotelian Society* 77: 195–223. https://doi.org/10.1111/1467-8349.00109

Thomasson, Amie. 1999. *Fiction and Metaphysics*. Cambridge University Press.

Thomasson, Amie. 2003. "Speaking of Fictional Characters". *Dialectica* 57: 205–223. DOI: 10.1111/dltc.2003.57.issue-2

Travis, Charles. 1994. "On Constraints of Generality". *Proceedings of the Aristotelian Society* 44, New Series, 44: 165–88 https://doi.org/10.1111/1746-8361.12199

Travis, Charles. 2017. "Views of my Fellows Thinking (Counting Thoughts)". *Dialectica* 71 (3): 337–378.

Tye, Michael. 1980. "In Defense of the Words 'Human Body'". *Philosophical Studies* 38 (2): 177–182.

Van Inwagen, Peter. 1980. "Philosophers and the Words 'Human Body'". In *Time and Cause. Philosophical Studies Series in Philosophy, vol 19*, edited by Van Inwagen, Peter, 283–299. Springer, Dordrecht. https://doi.org/10.1007/978-94-017-3528-5_16

Walton, Kendall. 1990. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press.

Walton, Kendall. 1993. "Metaphor and Prop Oriented Make-Believe". *European Journal of Philosophy* 1 (1): 39–57;

Walton, Kendall. 2015. *Other Shoes: Music, Metaphor, Empathy, Existence*. Oxford, New York: Oxford University Press.

White, Roger. 2011. "Throwing the Baby Out with the Ladder: On 'Therapeutic' Readings of Wittgenstein's Tractatus". In *Beyond the Tractatus Wars*, edited by Rupert Read, and Mathew Lavery, 22–65. London: Routledge.

Witherspoon, Edward. 2000. "Conceptions of Nonsense in Carnap and Wittgenstein". In *The New Wittgenstein*, edited by Alice Crary and Rupert Read, 315–349. Routledge.

Wittgenstein, Ludwig. 1918. *Logisch-philosophische Abhandlung*. Suhrkamp, 1963. English Translation: Ludwig Wittgenstein. *Tractatus Logico-Philosophicus*. Translated by Brian McGuiness, and David Pears. Taylor and Francis, 2002.

Wittgenstein, Ludwig. 1953. *Philosophische Untersuchungen* [*Philosophical Investigations*]. German-English Edition. Translated by Peter Hacker, Joachim Schulte, Elisabeth Anscombe, Oxford: Blackwell, 4th edition published in 2009.

Yablo, Stephen. 2001. "Go figure: A Path Through Fictionalism". *Midwest Studies in Philosophy* 25 (1): 72–102.

Contents

EDITORIALS

RESEARCH ARTICLES

DISCUSSION NOTES