

Volume XIV ♦ Issue 2

2023

Logos & Episteme

an international journal
of epistemology

**Romanian Academy
Iasi Branch**



**“Gheorghe Zane” Institute
for Economic and Social
Research**

Founding Editor

Teodor Dima (1939-2019)

Editorial Board

Editor-in-Chief

Eugen Huzum

Executive Editors

Vasile Pleșca

Cătălina-Daniela Răducu

Assistant Editors

Irina Frasin

Bogdan Ștefanachi

Ioan Alexandru Tofan

Web&Graphics

Codrin Dinu Vasiliu

Virgil-Constantin Fătu

Simona-Roxana Ulman

Contact address:

Institutul de Cercetări

Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

<http://logos-and-episteme.acadiasi.ro/>

https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme

Advisory Board

Frederick R. Adams

University of Delaware, USA

Scott F. Aikin

Vanderbilt University, USA

Daniel Andler

Université Paris-Sorbonne, Paris IV, France

Panayot Butchvarov

University of Iowa, USA

Mircea Dumitru

Universitatea din București, România

Sanford Goldberg

Northwestern University, Evanston, USA

Alvin I. Goldman

Rutgers, The State University of New Jersey, USA

Susan Haack

University of Miami, USA

Stephen Hetherington

The University of New South Wales, Sydney, Australia

Paul Humphreys

University of Virginia, USA

Jonathan L. Kvanvig

Baylor University, USA

Thierry Martin

Université de Franche-Comté, Besançon, France

Jürgen Mittelstrab

Universität Konstanz, Germany

Christian Möckel

Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot

Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins

Universidade de Lisboa, Portugal

Duncan Pritchard

University of Edinburgh, United Kingdom

Nicolas Rescher

University of Pittsburgh, USA

Rahman Shahid

Université Lille 3, France

Ernest Sosa

Rutgers University, USA

John F. Symons

University of Texas at El Paso, USA

TABLE OF CONTENTS

RESEARCH ARTICLES

| | |
|---|-----|
| Miguel López ASTORGA, Carnap Versus Popper: What Scientists Actually Do... | 131 |
| Ali HOSSEIN KHANI, Quine and First-Person Authority..... | 141 |
| Jimmy Alfonso LICON, Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing..... | 163 |
| Daniel RÖNNEDAL, Ought We to Believe the Truth and Nothing But the Truth? Two Arguments For the Wide Scope Version of the Truth Norm | 179 |

DISCUSSION NOTES/ DEBATE

| | |
|--|-----|
| Bálint BÉKEFI, Self-Favoring Theories and the Bias Argument..... | 199 |
| Howard SANKEY, Having a Hunch..... | 215 |

DISTRACTIONS

| | |
|--|-----|
| Jeremy FANTL, Matthew MCGRATH, Radical Knowledge Minimalism..... | 223 |
| Notes on the Contributors..... | 229 |
| <i>Logos and Episteme</i> . Aims and Scope..... | 231 |
| Notes to Contributors..... | 233 |

RESEARCH ARTICLES

CARNAP VERSUS POPPER: WHAT SCIENTISTS ACTUALLY DO

Miguel López ASTORGA

ABSTRACT: Carnap and Popper proposed ways scientists have to work. According to Carnap, they should look for confirmations for hypotheses. In Popper's view, what is important is to try to falsify hypotheses. Cognitive science seems to prove that, in real scientific research, both activities play a role. First, people attempt to confirm hypotheses. Second, they seek examples refuting those hypotheses. This paper is intended to show that the theory of mental models can describe the mental processes involved in both tasks: confirmation and falsification. It addresses the mental possibilities individuals consider in both cases. In addition, the paper reveals that, in accordance with both Carnap's framework and Popper's approach, both mental activities are related to conditional reasoning.

KEYWORDS: conditional reasoning, confirmation, falsifiability, mental models, possibilities

Introduction

In 1968, it was noted that, given a rule or hypothesis, people tend to confirm it (Wason 1968). Later, it was found that confirmation is only the activity people do at first. After it, individuals look for counterexamples, that is, they try to refute the rule or hypothesis (e.g., Tweney 1989; see also, e.g., Dunbar and Klahr 2012).

This discussion is important in epistemology and science. In philosophy of science there are different theoretical positions. Some of them claim that scientists need to confirm hypotheses (e.g., Carnap 1936, 1937). Other approaches ask for falsifying them (Popper 2002). Thus, what philosophy of science has proposed is what scientists must do. The relevance of works such as those in cognitive science (e.g., Dunbar and Klahr 2012; Tweney 1989; Wason 1968) is that they do not speak about what scientists should do, but about what they actually do.

The present paper will address the two tasks, confirmation and falsification, from the theory of mental models (e.g., Johnson-Laird 2023). Taking the models involved in each of these two activities, the paper will make two points. First, it will argue that the theory of mental models can describe the mental processes linked to confirmation and refutation. Second, it will also reveal that, if that description based on the theory of mental models holds, as proposed from different

approaches in philosophy of science (e.g., Carnap 1936, 1937; Popper 2002), scientific knowledge is related to conditional inferences.

The first section will be devoted to the findings about these issues in cognitive science. The next section will show how the theory of mental models can understand the confirmation process. It will propose that, within the theory, that process is a conditional reasoning process. The last section will do the same addressing the falsification process. It will present the way the theory of mental models can explain falsification. Then, it will indicate the reasons why, from the theory, that process is a conditional reasoning process too.

Hypotheses, Confirmation, and Refutation

Peter Wason (1968) was who noted that people tend to confirm hypotheses. Wason (1968) noted this fact by means of a reasoning task. The task presented a sequence with three numbers, for example, 2/4/6. Participants had to indicate more sequences. After that, they would get information on their sequences. In particular, they could know whether their sequences were coherent or incoherent with the rule the initial sequence followed. Almost all of participants thought that the rule provided that the second number in the sequence was the first number plus 2, and that the third number was the second number plus 2, the three numbers being even. The interesting finding was that participants gave examples of sequences confirming that hypothetical rule. That was not a good strategy, since the right rule was different: the next number in the sequence is higher than the previous one (see also, e.g., Dunbar and Klahr 2012).

The wrong rule participants supposed can be expressed as (1).

- (1) $N_i/N_j/N_k$, where N_i , N_j , and N_k are even numbers, $N_j = N_i + 2$, and $N_k = N_j + 2 = N_i + 4$

So, what participants did is to offer sequences confirming (1) such as 6/8/10 or 26/28/30.

However, as said, the true rule was simpler. It was (2).

- (2) $N_i/N_j/N_k$, where N_i , N_j , N_k are numbers, and $N_k > N_j > N_i$

Works such as that of Tweney (1989) changed this idea, or rather complemented it. Those works revealed that confirmation is just the first phase in the mental process by means of which people review hypotheses. When individuals consider confirmation to be enough, they start to think about cases refuting the hypothesis (see also, e.g., Dunbar and Klahr 2012).

If the example of Wason's (1968) task is taken into account, the discovery of the second phase would imply to propose sequences such as 2/4/5 or 2/4/7 in that

new phase. These sequences would violate (1), but they would be consistent with, for instance, (2).

From the point of view of philosophy of science, it can be said that individuals first act as philosophers such as Carnap (1936, 1937) point out. After that, they do what Popper (2002) claims. As mentioned, what is important about this is that both Carnap (1936, 1937) and Popper (2002) only refer to what scientists should do, not to what they really do. What scientists actually do can be discovered from works such as those analyzed in this section.

The next section is devoted to the first phase in the review of hypotheses, that is, to confirmation. It is dealt with from the theory of mental models. Inter alia, it is shown that, from this theory, confirmation processes imply conditional reasoning.

Confirmation and the Theory of Mental Models

If it is assumed that s_1, s_2, \dots, s_n are sequences with numbers, and that

$Px =_{df} x$ has three numbers

Then,

$Ps_1 =_{df} s_1$ has three numbers

$Ps_2 =_{df} s_2$ has three numbers

.

.

.

$Ps_n =_{df} s_n$ has three numbers

If it is added that

$Qx =_{df} x$ is consistent with (1)

It can be derived that

$Qs_1 =_{df} s_1$ is consistent with (1)

$Qs_2 =_{df} s_2$ is consistent with (1)

.

.

.

$Qs_n =_{df} s_n$ is consistent with (1)

In this way, when individuals search for confirmations for (1), what they do is to search for cases such as those expressed in (3).

(3) $(Ps_1 \ \& \ Qs_1) \ \& \ (Ps_2 \ \& \ Qs_2) \ \& \dots \& \ (Ps_n \ \& \ Qs_n)$

Conjunction (3) indicates that sequences with three numbers coherent with (1) are possible. (3) gives n examples. Accordingly, (3) allows assuming (4).

(4) PS ($Px \ \& \ Qx$)

Where 'PS' means that what follows between brackets is possible, in the sense 'possible' has in natural language (not in the sense it has in modal logic).

Although (3) offers n examples, just one would suffice to affirm (4). But what is most important is that, contrary to what modal logic requires, people often think that, if something is possible, its denial is also possible (e.g., Espino, Byrne, and Johnson-Laird 2020). Hence, if (4) is a possibility, (5) is a possibility as well.

(5) PS [$\text{Not}-(Px \ \& \ Qx)$]

If (5) is a possibility, then (6), (7), and (8) are possibilities too.

(6) PS ($Px \ \& \ \text{Not}-Qx$)

(7) PS ($\text{Not}-Px \ \& \ Qx$)

(8) PS ($\text{Not}-Px \ \& \ \text{Not}-Qx$)

Possibilities (7) and (8) are not problematic. In them, x does not have three numbers. They refer to cases in which x does not fulfill the first condition (i.e., P). That means that x does not match the structure of the sequences in Wason's (1968) task. For that reason, (7) and (8) cannot falsify (1). The only possibility that would refute (1) would be (6). The latter possibility would indicate that x has three numbers and, nevertheless, it is inconsistent with (1). Thereby, it can be said that there is a 'conjunction of possibilities' (which is the expression used in the theory of mental models; e.g., Khemlani, Hinterecker, and Johnson-Laird 2017) representing the kinds of scenarios in which (1) can be the case. That is conjunction of possibilities (9).

(9) PS ($Px \ \& \ Qx$) & PS ($\text{Not}-Px \ \& \ Qx$) & PS ($\text{Not}-Px \ \& \ \text{Not}-Qx$)

This conjunction of possibilities is the conjunction that, following the theory of mental models, corresponds to the conditional (e.g., Quelhas, Rasga, and Johnson-Laird 2017). According to the theory of mental models, sentences are related to the models in which they are true. A possibility expresses a model of that type (see also, e.g., Johnson-Laird and Ragni 2019). A conjunction of possibilities such as (9) corresponds to a sentence such as (10) (see also, e.g., Goodwin and Johnson-Laird 2018).

(10) If Px then Qx

Several points are interesting about this. The theory of mental models is a psychological theory, not a logical one. P and Q do not have the properties predicates have in calculi such as first-order predicate calculus. (10) is not a well-formed formula in that calculus. The theory of mental models only tries to account for what people do in daily life. But in daily life, individuals do not always deduce that the possibilities that can be linked to a sentence such as (10) with thematic content are those in (9) (see also, e.g., Byrne and Johnson-Laird 2020). For example, (7) and (8) are often ignored (see also, e.g., Byrne and Johnson-Laird 2009). In this way, it is important that the theory deems (7) and (8) as presuppositions. This is because, as shown below, they are valid possibilities both if (10) is true and if (10) is false (see also, e.g., López-Astorga, Ragni, and Johnson-Laird 2022). So, to look for cases of (4) is, from this point of view, to look for cases that actually confirm a conditional such as (10).

Therefore, the theory of mental models can describe the confirmation mental processes. It can be thought that those processes happen in the first phase when a scientific hypothesis is addressed. The next section tries to argue the same with regard to the falsification mental processes, that is, the processes occurring in the second phase of analysis of scientific hypotheses.

Refutation and the Theory of Mental Models

In a similar way as in the previous mental process, when people try to falsify, they search for cases such as those in (11).

$$(11) (P_{S1} \& \text{Not-}Q_{S1}) \& (P_{S2} \& \text{Not-}Q_{S2}) \& \dots \& (P_{Sn} \& \text{Not-}Q_{Sn})$$

If the cases included in (11) were found, that would show that there are sequences with three numbers violating (1), and that hence the latter rule is false. For the example considered in this paper, that is, that of Wason's (1968) task, the cases in (11) could be, for instance, cases inconsistent with (1) but compatible with (2). (11) would inform n cases revealing that (1) does not hold. Thereby, (12) could be concluded from (11).

$$(12) PS (Px \& \text{Not-}Qx)$$

As mentioned above for confirmation, although (11) gives n examples, one example would be enough to assume (12). As also said, the theory of mental models supposes the idea that if something is possible, its opposite is possible too. This idea can also be applied to (12). So, possibility (12) implies possibility (13).

$$(13) PS [\text{Not-}(Px \& \text{Not-}Qx)]$$

Possibility (13) leads to possibilities (4), (7), and (8). This is because (4), (7), and (8) are the possibilities that, in principle, appear to make (12) false. Nonetheless, as in the case of (4) and (5), there are two possibilities that are not a problem for (12). Again, they are those in which the first condition is not fulfilled, that is, those including Not-P, that is, (7) and (8). For the same reasons as in confirmation (7) and (8) cannot falsify (1), they cannot falsify (12) either. The only controversial possibility confirming (1) is (4) now. As mentioned, (4) expresses that x has three numbers and is consistent with (1). The conjunction of possibilities for falsification is (14).

(14) $PS(Px \ \& \ \text{Not-}Qx) \ \& \ PS(\text{Not-}Px \ \& \ \text{Not-}Qx) \ \& \ PS(\text{Not-}Px \ \& \ Qx)$

Conjunction of possibilities (14) is the conjunction the theory of mental models assigns to the negation of the conditional. This is one more difference between the theory of mental models and standard logics. Standard logics tend to understand the conditional as Philo of Megara did (e.g., Bocheński 1963; O'Toole & Jennings 2004). As it is well known, from Philo's view, the negation of a conditional such as (10) leads to a conjunction such as (15).

(15) $Px \ \& \ \text{Not-}Qx$

Nevertheless, the theory of mental models interprets the conditional based on what empirical experiments reveal. According to the theory, there are experiments (e.g., those in Khemlani, Orenes, and Johnson-Laird 2014) showing the way people actually deny conditionals. Thus, following the results of those experiments, the theory of mental models considers (16) to be the negation of (10) (see also, e.g., López-Astorga et al. 2022).

(16) $\text{If } Px \text{ then } \text{Not-}Qx$

Hence, from what was said about the conditional within the theory of mental models in the previous section, it can be stated that the theory provides that the negation of a conditional corresponds to a conjunction of possibilities such as (14). In (14), the second and third possibilities, that is, respectively, (8) and (7), keep being presuppositions. They can be accepted whether the conditional is true or false. So, they can be accepted both for (10) and for (16).

All of this means that to look for cases of (12) is to look for cases confirming (16) and, therefore, refuting (10). Presuppositions are valid whether or not the conditional holds. Accordingly, they are not the key here.

In summary, the theory of mental models can also describe the mental processes that happen in the second phase when analyzing a hypothesis. This is because the theory can describe the refutation mental processes. In these processes,

reasoning is related to the conditional too. This is not a problem. It is rather a strength of the theory of mental models. The reasons for that are two. On the one hand, in the reduction sentences Carnap (1936, 1937) proposes, that is, in the sentences that, following Carnap (1936, 1937), should be used in confirmation, the conditional structure is essential. On the other hand, Popper (2002) understands the falsification processes as the application of *Modus Tollendo Tollens*. One of the premises in the latter rule is a conditional.

Conclusions

Philosophers of science such as Carnap and Popper tried to show the manner scientists should work to accept or reject hypotheses. The former is a representative of a school of thought claiming confirmation as the basic task. On the contrary, Popper argued that the scientific task needs to be based on refutation.

Proposals such as these are presented as what scientists must do. But cognitive science has revealed how people actually act when they try hypotheses. From works such as those of Tweney or Wason, one might infer that there are two phases of analysis. First, individuals attempt to confirm hypotheses. Then they search for refutations of those very hypotheses (see also, e.g., Dunbar and Klahr 2012).

What the present paper shows is that the theory of mental models can explain how both of those processes happen in the human mind. Confirmation is understood as the quest for possible cases in which both the antecedent and the consequent of a conditional hold. The other possibilities in which that conditional can be the case are not addressed because they are presuppositions. They refer to scenarios that are also possible if the conditional is false. As far as falsification is concerned, the theory interprets it as the process seeking possible circumstances in which the antecedent of a conditional is true and the consequent of that very conditional is false. Again, the other possible situations in which the conditional can be false are not dealt with. They are presuppositions as well. In fact, they are the same presuppositions as those corresponding to the conditional when it is true.

On the one hand, these results are not far from logical and philosophical conceptions such as those of Carnap or Popper. Both of them deemed the processes of analysis of scientific hypotheses as processes in which the conditional is involved. The structure of Carnap's reduction sentences is conditional. In Popper's view, refutation reproduces the application of a logical rule containing a conditional: *Modus Tollendo Tollens*. On the other hand, these results can also be understood as evidence in favor of the theory of mental models. The latter theory is not intended to account for what the human mind should do, but what it

actually does. The present paper has tried to argue that the theory of mental models is able to describe what scientists actually do.¹

References

- Bocheński, Józef Maria. 1963. *Ancient Formal Logic*. Amsterdam: North-Holland.
- Byrne, Ruth M. J., and Philip N. Johnson-Laird. 2009. "If and the problems of conditional reasoning." *Trends in Cognitive Science* 13 (7): 282-287. <https://doi.org/10.1016/j.tics.2009.04.003>
- Byrne, Ruth M. J., and Philip N. Johnson-Laird. 2020. "If and or: Real and counterfactual possibilities in their truth and probability." *Journal of Experimental Psychology: Learning, Memory, and Cognition* 46 (4): 760-780. <https://doi.org/10.1037/xlm0000756>
- Carnap, Rudolf. 1936. "Testability and meaning." *Philosophy of Science* 3 (4): 419-471. <https://doi.org/10.1086/286432>
- Carnap, Rudolf. 1937. "Testability and meaning – Continued." *Philosophy of Science* 4 (1): 1-40. <https://doi.org/10.1086/286443>
- Dunbar, Kevin N., and David Klahr. 2012. "Scientific thinking and Reasoning." In *The Oxford Handbook of Thinking and Reasoning*, edited by Keith J. Holyoak and Robert G. Morrison, 701-18. New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199734689.013.0035>
- Espino, Orlando, Ruth M. J. Byrne, and Philip N. Johnson-Laird. 2020. "Possibilities and the parallel meanings of factual and counterfactual conditionals." *Memory & Cognition* 48: 1263-1280. <https://doi.org/10.3758/s13421-020-01040-6>
- Goodwin, Geoffrey P. and Philip N. Johnson-Laird. 2018. "The truth of conditional assertions." *Cognitive Science* 42(8): 2502-2533. <https://doi.org/10.1111/cogs.12666>
- Johnson-Laird, Philip N. 2023. "Possibilities and human reasoning." *Possibilities Studies & Society*. <https://doi.org/10.1177/27538699231152731>
- Johnson-Laird, Philip N., and Marco Ragni. 2019. "Possibilities as the foundation of reasoning." *Cognition* 193. <https://doi.org/10.1016/j.cognition.2019.04.019>
- Khemlani, Sangeet, Thomas Hinterecker, and Philip N. Johnson-Laird. 2017. "The provenance of modal inference." In *Computational Foundations of*

¹ **Acknowledgments:** PIA Ciencias Cognitivas, Centro de Investigación en Ciencias Cognitivas, Instituto de Estudios Humanísticos, Universidad de Talca; Fondo Fondecyt de Continuidad para Investigadores Senior, código FCSN2102, Universidad de Talca; Fondecyt (Programa de Equipamiento Científico y Tecnológico) 2019, código EQM190153.

- Cognition*, edited by Glenn Gunzelmann, Andrew Howes, Thora Tenbrink, and Eddy J. Davelaar, 663-68. Austin: Cognitive Science Society.
- Khemlani, Sangeet, Isabel Orenes, and Philip N. Johnson-Laird. 2014. "The negation of conjunctions, conditionals, and disjunctions." *Acta Psychologica* 151: 1-7. <https://doi.org/10.1016/j.actpsy.2014.05.004>
- López-Astorga, Miguel, Marco Ragni, and Philip N. Johnson-Laird. 2022. The probability of conditionals: A review. *Psychonomic Bulletin & Review* 29: 1-20. <https://doi.org/10.3758/s13423-021-01938-5>
- O'Toole, Robert R., and Raymond E. Jennings. 2004. "The Megarians and the Stoics." In *Handbook of the History of Logic, Volume 1. Greek, Indian and Arabic Logic*, edited by Dov M. Gabbay and John Woods, 397-522. Amsterdam: Elsevier. [https://doi.org/10.1016/S1874-5857\(04\)80008-6](https://doi.org/10.1016/S1874-5857(04)80008-6)
- Popper, Karl. 2002. *The Logic of Scientific Discovery*. London: Routledge. <https://doi.org/10.4324/9780203994627>
- Quelhas, Ana Cristina, Célia Rasga, and Philip N. Johnson-Laird. 2017. "A priori true and false conditionals." *Cognitive Science* 41 (55): 1003-1030. <https://doi.org/10.1111/cogs.12479>
- Tweney, Ryan D. 1989. "A framework for the cognitive psychology of science." In *Psychology of Science: Contributions to Metascience*, edited by Barry Gholson, Arthur Houts, Robert A. Neimeyer, and William Shadish, 342-66. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CBO9781139173667.018>
- Wason, Peter C. 1968. "Reasoning about a rule." *Quarterly Journal of Experimental Psychology* 20 (3): 273-281. <https://doi.org/10.1080/14640746808400161>

QUINE AND FIRST-PERSON AUTHORITY

Ali HOSSEIN KHANI

ABSTRACT: Blackburn and Searle have argued that Quine's thesis of the indeterminacy of translation results in a denial of the sort of first-person authority that we commonly concede we have over our mental and semantical content. For, the indeterminacy thesis implies that there is no determinate meaning to know at all. And, according to Quine, the indeterminacy holds at home too. For Blackburn, Quine must constrain the domain of indeterminacy to the case of translation only. Searle believes that Quine has no other choice but to give up on his behaviorism. Hylton, however, has attempted to defend Quine against these objections, by arguing that Quine's naturalistic claim that speaking a language is nothing but possessing certain dispositions to act in specific ways would enable him to accommodate first-person authority. I will argue that the objections from Blackburn and Searle, as well as Hylton's solution, are all problematic when seen from within Quine's philosophy. I will introduce a sort of Strawsonian-Wittgensteinian conception of first-person authority and offer that it would be more than compatible with Quine's naturalistic philosophy.

KEYWORDS: indeterminacy of translation, first-person authority, Quine, Blackburn, Hylton, Strawson

1. Introduction

What would Quine say of the first-personal authority that we strongly concede we have over the meaning of our own utterances? To put the question in terms of the attributions of meaning (and mental content) to others and ourselves, could Quine offer an explanation of the asymmetry between self-ascriptions of meaning and ascriptions of meaning to others' utterances? Very few philosophers have attempted to answer these questions. I think everyone agrees that any plausible answer must take into account Quine's central remarks on naturalized epistemology, criteria of individuation, stimulus meaning, and the indeterminacy-underdetermination distinction. My aim in this paper is not to go through all such topics in detail; they have been well discussed in the extensive literature on Quine.¹ I will rather assume familiarity with most of them and attempt to unpack

¹ See, e.g., Gibson (1986; 2004), Hossein Khani (2018; 2021a; 2023), Hylton (2007), Kemp (2006), Soames (1999), Orenstein (2002). and Verhaegh (2018).

the problem which seems to appear within Quine's naturalistic philosophy with regard to the existence of such an asymmetry in meaning-attributions. Quine does not explicitly address such an issue. My chief goal is to envisage what can be extracted about this matter from his remarks on meaning and translation.

One answer to the above related questions is that Quine really has no other choice but to deny that we can say anything philosophically interesting about any asymmetry in attributions of a *determinate* meaning: the indeterminacy of translation has it as its *sceptical conclusion* that there is no fact as to what the correct translation of a speaker's utterances is and if all that can be said about such a content is confined to what we can extract from the Quinean reflections on radical translation and its accompanied indeterminacy, we have no other way but to concede that there is nothing uniquely determinate about meaning to know at all. If there is no such thing as a determinate meaning, there is no such thing as knowing it, directly or indirectly, inferentially or otherwise. Nor is there any determinate meaning to attribute to the utterances of others and ourselves. Quine once declared that "a sentence has a meaning, people thought, and another sentence is its translation if it has the same meaning. This, we see, will not do" (1987, 8).²

Some critics of Quine have criticized him on the basis of this negative conclusion of his arguments for the indeterminacy of translation. Since Quine's indeterminacy thesis results in a denial of the existence of any fact about determinate meanings – which if we were supposed to have authoritative knowledge of, it would be of them – he thereby fails to accommodate an essential asymmetry between the subject (the speaker) and others (the translators) with regard to knowledge of what the subject means by her expressions, or at least with regard to the essential difference there is between the basis for self-ascriptions of meaning and that of the ascriptions of meaning to the speaker by the translator. Up to some point, our concern may merely (but legitimately) be to answer the question how an alien language can be translated into our own; but, as the objection goes, it is unacceptable to conclude from an answer to that question that

² See also Quine (1960, 73; 1968, 275; 1987, 10; 1995a, 75-76). This is sometimes taken to be the view of the "early" Quine. By "early" I mean his pre-1975 works on the topic, such as Quine (1960; 1968; 1969b; 1970). By the "later" Quine, I mean the period of his works on the topic that starts with Quine (1975), and then Quine (1987; 1981) and continues to his later works such as Quine (1990a; 1990c; 1995a). The main difference between them is the way Quine treats the indeterminacy of translation thesis, i.e., as a "conjecture" (1986, 728), in which underdetermination plays no significant role. This difference is not the concern of this paper because Quine never gives up on the claim that meaning *is* indeterminate. He remains a non-factualist about such traditional, or fine-grained, meanings and this is what matters in this paper.

this is *all* we can do about meaning and knowledge of it. We cannot genuinely hold that there really is no difference whatsoever between the speaker's self-attributions of meaning and the speaker's attributions of meaning to others' utterances. Failing to appreciate the existence of such an asymmetry misses a philosophically vital distinction, the distinction between the first-person and the second-person, or one may even add, the distinction between the subjective and the objective, the self and the world. Quine does not seem to be on board with such an outcome either. If so, appreciating the existence of such an asymmetry would inevitably force him to offer an explanation of, or at least a plausible story about, this phenomenon which seems to have been absent in Quine's works.

Blackburn (1984) and Searle (1987) – and more recently Glock (2003, 201–207) – are among those critics who think that Quine's thesis of the indeterminacy of translation leads to such a highly counterintuitive consequence, i.e., a denial of the existence of such asymmetries. Hylton (2007; 1990/91) disagrees and attempts to show that Quine does not need to give up on the existence of such a phenomenon. In what follows, I examine the objections from Blackburn and Searle, as well as Hylton's defense of Quine. The extent to which I agree with these philosophers is only that Quine is required to say something constructive about such asymmetries. I will, however, argue that in one way or another the objections from Blackburn and Searle, as well as Hylton's defense, all miss certain central points in Quine's naturalistic philosophy. Among other things, they treat Quine as if he is on board with the idea that meaning can be divorced from translation, that indeterminacy does not arise at home, that the pragmatic criteria, such as smoothness of dialogue, can be treated as somehow *determining* facts about meanings. None of these can be done in Quine's philosophical framework. I will then offer a sort of Strawsonian-Wittgensteinian account, the most important feature of which is its neutrality to the questions about the metaphysical status of meaning and which I think Quine could employ without jeopardizing the consequences of his arguments for the indeterminacy thesis. I begin with an important distinction with regard to the notion of meaning.

2. Stimulus Meaning vs. Fine-Grained Meaning

Quine's arguments for the indeterminacy of translation aim to establish the sceptical conclusion that there is no fact as to what the correct translation of an utterance is and thus what a speaker, *as a matter of fact*, means by her words.³

³ More particularly, his "Argument from Below" and the "Argument from Above". See, e.g., Quine (1970, 183).

Radical translation focuses on the process of translating an entirely unknown language spoken by a native speaker. The native assents to “Gavagai” in the presence of (the stimulations of) a rabbit in enough cases and this evidence allegedly convinces the translator to translate this one-word sentence into “Lo, a rabbit,” on the basis of the fact that, in a similar situation, she would assent to “Lo, a rabbit.” Quine argues that there are always rival translations of the native’s sentences that are mutually incompatible with each other but are all compatible with all possible (allowed-by-Quine) facts of the matter – that is, the physical facts in general and facts about the native’s verbal dispositions to assent to or dissent from certain sentences on specific occasions. The famous examples of such alternative translations are “Lo, an undetached rabbit-part,” “Lo, another manifestation of rabbithood,” and so forth.⁴

The second important claim from Quine is that he does not see the indeterminacy as solely emerging in the process of translating a radically unknown language; rather, as he emphasizes, “radical translation begins at home” (1969, 46) and thus, the indeterminacy manifests itself at home too: “The problem at home differs none from radical translation ordinarily so called” (Quine 1969, 47) so that “[r]eference would seem... to become nonsense not just in radical translation but at home” (1969, 48). One reason for this is that, for Quine, “there is no entity without identity; [thus] no meaning without sameness of meaning” (1995a, 75-76).⁵ Quine makes it even easier to see why translation and indeterminacy both hold at home:

I have directed my indeterminacy thesis on a radically exotic language for the sake of plausibility, but in principle it applies even to the home language. For given the rival manuals of translation between Jungle and English, we can translate English perversely into English by translating it into Jungle by one manual and then back by the other. (1990a, 48)

Meaning and translation are indeterminate everywhere. Consequently, Quine does not, and cannot, grant the translator a language in which meanings are uniquely determinate.

We can also detect a second reason for such a claim: if meanings are determinate for the translator in his own language, they are determinate for the native speaker in her own language as well. The consequence of this is that there *are* thereby facts about determinate meanings and all the translator is doing turns into the task of using the available evidence to capture such facts. This means that the indeterminacy problem becomes a mere *epistemological* problem, rather than a *sceptical* problem with the *ontological* consequence that there are no such things

⁴ See, e.g., Quine (1960, 52-53, 71-72; 1969b, 30-34; 1987, 7; 1990a, 45; 1995a, 71-72).

⁵ See especially Quine (1948).

as determinate meanings. The indeterminacy problem now turns into a very similar doctrine like that of the underdetermination of translation by evidence, which is essentially an epistemological problem for Quine. I will come back to this problem later. Thus, it is vital to note that the consequence of Quine's arguments for the indeterminacy of translation is that reference is in general inscrutable and meaning is generally indeterminate so that there is no fact as to what someone means by her words, whether the language is radically unknown to us or it is our own familiar language.

The indeterminacy of translation undermines our traditional conception of meaning, and this is the reason why Quine states that "[w]hat I have challenged is... an ill-conceived notion within traditional semantics" (1987, 10). Quine's attack is on the existence of *fine-grained meanings*. According to Quine, "Gavagai" and "Lo, a rabbit" have the same "stimulus meaning." He defines the stimulus meaning of a sentence, for a given speaker, as (the ordered pair of) all the stimulations that would prompt the speaker to assent to, and dissent from, that sentence.⁶ We can then explain an expression's fine-grained, or traditional, meaning as follows: two expressions possess different fine-grained meanings if they have the same stimulus meaning but differ in the sort of unique meaning that we (traditionally) expect them to possess.⁷ In this sense, the competing translation manuals are alike with regard to the sentences' stimulus meaning but differ regarding the alleged fine-grained meanings that we expect them to possess – after all, "Lo, a rabbit" and "Lo, an undetached rabbit-part" seem to have different fine-grained meanings. Quine's claim is that there would always be a slack between stimulus meanings and fine-grained meanings, a gap which can never be filled because facts about stimulus meaning are too weak to constitute one unique fact about fine-grained meanings. We can never ascend from stimulus meanings to fine-grained ones.

The difference between Quine, who is a non-factualist about fine-grained meanings, and a factualist about meanings, e.g., a non-reductionist, is that for the non-reductionist there *are* facts about fine-grained meanings, though this claim would not necessarily stand against the idea that (possible) alternative meaning facts could exist: "Green" means *green* but it could have meant something else. For our non-reductionist too, there can be endless possible meaning facts with regard to each expression of language. Nonetheless, the crucial difference between Quine and the non-reductionist is that, for the latter, facts about fine-grained meanings are primitive so that the rest of all possible meaning facts are automatically *out of*

⁶ See Quine (1960, 32-33).

⁷ For a recent discussion of this, see Hossein Khani (2018; 2023).

the set of all actual facts about meaning.⁸ For Quine, however, there is no set of *actual* facts about fine-grained meanings separate from the set of *possible* facts about them. The *actual* facts about stimulus meaning cannot constitute one unique set of actual facts about fine-grained meanings and thereby rule out the rest. Consequently, no plausible account of self-knowledge seems to be in peace with the idea that the speaker directly and non-inferentially knows an *indefinite* number of meanings of one expression.⁹ This makes it difficult to provide a Quinean account of the asymmetry in question *without* abandoning his naturalistic notion of stimulus meaning. I think the way out of this problem is to look for an account that remains *neutral* to the metaphysical status of meaning.

One way to do that, one may suggest, is to appeal to the notion of *agreement in use*. But we have just postponed the problem: How should we treat such facts about agreement in use? Is there an agreement in using a word *because* certain facts about its fine-grained meaning obtain or is there such an agreement simply *because* it is simply an empirical fact that members of a speech-community are all trained or conditioned to be disposed (or blindly inclined) to respond to the world and to each other's responses in one way rather than another? Quine's naturalism seems to deny the existence of the first sort of facts and only allow for the second, by confining the study of meaning to the study of the link between our best theory of reference (and meaning) and the flow of relevant evidence, which itself is constrained by Quine's physicalism, according to which "[n]othing happens in the world... without some redistribution of microphysical states" (Quine 1981, 98).¹⁰ No change in meaning is possible, unless there is some physical change of one sort somewhere, which is, in the case of translation, a change in the speaker's behavioral dispositions to assent and dissent. We can say that Quine's physicalism *manifests* itself as a sort of behaviorism in the case of translation: "there is nothing in linguistic meaning... beyond what is to be gleaned from overt behaviour in observable circumstances" (1987, 5). The totality of such behavioral facts, however, fails to constitute one unique fact about an expression's fine-grained meaning.

⁸ For instance, Wright's non-reductionist judgement-dependent account of meaning and intention has such a characteristic. See Wright (1992; 2001).

⁹ This is a claim different from the (plausible) claim (sustained by a factualist) that the unique meaning of an expression, which one grasps, can have *an indefinite number of normative consequences*, that is, that meanings determine the correct use of expressions for a potentially infinite number of occasions.

¹⁰ For more on Quine's physicalism, see Føllesdal (2014), Glock (2003), Harman (2014), Hookway (1988, 212), Hossein Khani (2018; 2021a), Hylton (2007, 1982), Kemp (2006), and Kirk (1986, 1969).

Therefore, if we wish to bring in the notion of agreement in use, a desired Quinean explanation of first-person authority must be able to work with facts about agreement in use viewed only as a sort of primitive facts about the speech-community, which the speaker is a member of, i.e., as not being constituted by any deeper fact about fine-grained meanings and as not capable of constituting any fact about fine-grained meanings. This will be the main target of this paper to find such an account. I believe a workable account is Strawson's (sort of Wittgensteinian) account. But let me first review the objections from Blackburn and Searle and the problems with their solutions.

3. The Objections from Blackburn and Searle

According to Searle, "[i]f the argument [from indeterminacy] is valid, then it must have the result that there isn't any difference *for me* between *meaning* rabbit or rabbit stage, and that has the further result that there isn't any difference for me between *referring* to a rabbit and referring to a rabbit stage" (1987, 130). He thinks that this implausible conclusion is a consequence of Quine's broad application of behaviorism and forms a *reductio* for Quine's behaviorism.¹¹ The idea behind this objection is that indeterminacy cannot arise at home because if it does, it amounts to a denial of first-person authority. A similar point is made by Glock: "I deny that radical translation and hence indeterminacy starts at home" (2003, 201). Glock is sympathetic with Searle's claim that "[w]e know from our own case, from the first-person case, that behaviorism is wrong, because we know that our own mental phenomena are not equivalent to dispositions to behavior" (Searle 1987, 136).¹² For Searle and Glock, "Quine's indeterminacy argument... denies the existence of distinctions that we know from our own case are valid" (Searle 1987, 137). Glock, like Searle, thinks that first-person vs. second-person asymmetries form a constitutive part of our semantic concepts and that denying it is to be taken to form a *reductio* for any view implying such a denial. The consequence of Quine's indeterminacy thesis, when so broadly applied, would be to conclude that "he would then no longer be talking about meaning as commonly understood" (Glock 2003, 206). But did Quine ever claim that he is talking about meaning as *commonly understood*? It does not seem so as he declared that his aim is to undermine meaning as commonly (traditionally) understood.¹³

¹¹ See Searle (1987, 136-137). For a discussion of this, see also Føllesdal (1990).

¹² See Glock (2003, 206).

¹³ Though, I will use "commonly" in the Wittgensteinian way later, which departs from "traditionally" as Quine used it.

Searle's objection pictures Quine's view as if Quine can abandon his behaviorism without any harm. But we saw that Quine's behaviorism stems from his general naturalistic view: more particularly, from his physicalism. For him, the naturalistically salient facts in the case of meaning are facts about the speaker's dispositions to assent and dissent, or basically behavioral facts. If they fail to fix facts about fine-grained meanings – and Quine thinks they fail – the conclusion would be that there is simply no fact about such meanings at all. Quine's behaviorism is the view that the naturalistically worth noting facts, in the case of translation, are behavioral. Otherwise, it would become vague what sort of facts must be taken into account when our concern is the case of translation. As Quine himself puts it, "where I have insisted on behaviorism is in linguistics" (1990b, 291). By demanding Quine to give up on his behaviorism here, we are asking Quine to allow for facts *over and above* the physical (i.e., behavioral) facts in the case of translation. What would be such alleged facts, however? Obviously, they cannot be facts about fine-grained meanings: Quine's arguments for the indeterminacy denied their existence. The first problem with Searle's objection is thus that he seems to overlook the fact that his demand, that Quine retreats from such a special sort of behaviorism, has serious destructive consequences for Quine's philosophy. There are, however, more serious problems with Searle's objection, as well as his solution. Let me consider these problems after introducing Blackburn's objection and proposed solution.

Blackburn thinks that if Quine's indeterminacy thesis applies to the case of the speaker *herself*, it would lead to a highly unacceptable conclusion. As he puts it, "[b]ut after all, I know what I mean when I say that rabbits are good to eat – I know what I refer to and what I say. Indeterminacy may afflict the bleak physicalistic outsider looking at me, but to me and to my fellow-speakers there is no shadow of indeterminacy to be seen" (1984, 281). Blackburn's claim is that, in my own case, no indeterminacy *can* emerge: there can be no shadow of indeterminacy *in the case of the translator using her own language*. This means that there is then a sense in which speakers *know* what they mean by their *own* words, as if we can hold onto meaning *without* translation, as if translation can be divorced from meaning in such a way that the translator can be credited with knowledge of the determinate meanings of the terms of her own language. Like Searle, Blackburn too sketches an outline of a possible solution. What he seems to be suggesting is that some *further* constraint should be imposed on Quine's physicalism if Quine is to have any chance to rescue his account from turning into a highly counterintuitive one. We do not need to go through the details of what such constraints can be because it is already clear that a problem similar to the one

with Searle's objection arises in the case of Blackburn's suggestion too. For, imposing extra conditions on Quine's naturalism and its resulted physicalism is already considered as illegitimate by and from within Quine's naturalism, according to which philosophy is not "an *a priori* propaedeutic or groundwork for science, but as continuous with science" (1969, 126) so that when you "see language naturalistically... you have to see the notion of likeness of meaning... as simply nonsense" (1969, 30). We can put the problem with Blackburn's suggestion in the form of a dilemma for him: either the constraints on Quine's physicalism which he has in mind are supposed to allow for the existence of determinate (fine-grained) meanings in the language of the speaker herself, in which case Quine has already rejected any such possibility, or the alleged constraints are not supposed to work in that way, in which case it would not be clear in what sense they are supposed to help deal with the present problem at all. For, if the latter is the case, the constraints would inevitably leave the sceptical consequences of the indeterminacy thesis untouched. We are then back to where we started. Let me now introduce what I take to be more serious problems with both Searle's and Blackburn's objections. I concentrate on three related problems.

Searle and Blackburn both claim that indeterminacy does not arise at home. This is a highly problematic claim within Quine's philosophy, for two reasons. The first is textual, which I mentioned earlier in this paper. Quine insists that indeterminacy holds in the speaker's own language as well: "radical translation begins at home" (1969, 46) and thus the indeterminacy manifests itself at home too. Again, "[t]he problem at home differs none from radical translation ordinarily so called" (1969, 47) so that "[r]eference would seem... to become nonsense not just in radical translation but at home" (1969, 48). Indeterminacy *does* arise at home for the reason which I briefly mentioned before and which puts forward the second problem with Searle's and Blackburn's suggestions: if we think that translation and indeterminacy do not hold at home, we are misreading the indeterminacy thesis as an *epistemological* problem. If meanings *are* fixed for the translator and within the translator's own language, there are then *facts* about fine-grained meanings. But then just like the case of the translator, in the case of the native's language too meanings would be determined – the native is also a speaker and a translator at the same time. There are thereby meaning facts. In this case, the only remaining problem would be that the behavioral evidence would not suffice to pick out one unique translation among the rivals. Nonetheless, that would not imply that there are *no meaning facts*.¹⁴ This is to make the same kind of mistake

¹⁴ Compare this to Quine's metaphor of the "Myth of a Museum" in Quine (1969, 27-30; 1981, 41).

that, according to Quine, Chomsky (1968) makes with regard to the indeterminacy thesis. What was the mistake?

For Quine, translation is underdetermined by all possible evidence so that “the totality of possible observations of verbal behavior... is compatible with systems of analytical hypotheses of translation that are incompatible with one another” (1968, 274-275). But Quine’s claim is that there is an essential difference between underdetermination and indeterminacy: the latter “is parallel but additional” (1968, 275). Chomsky thinks that translation’s being underdetermined by behavioral evidence is all that is philosophically salient about translation: “It is, to be sure, undeniable that if a system of ‘analytical hypotheses’ goes beyond evidence then it is possible to conceive alternatives compatible with the evidence ... Thus the situation in the case of language ... is, in this respect, no different from the case of physics” (Chomsky 1968, 61). For Quine, however, Chomsky fails to appreciate the indeterminacy-underdetermination distinction and this is the point that “Chomsky did not dismiss... He missed it” (Quine 1968, 67). Not only are the translation manuals underdetermined by all possible evidence, but they are also indeterminate because “there is no fact of the matter [about correct translation] even to *within* the acknowledged under-determination of a theory of nature” (Quine 1968, 275). Translation withstands the whole truth about the world because it remains underdetermined even when all of the physical facts are fixed *within* our favored theory of the world. This means that the totality of all physical facts fails to determine facts about correct translation. There is thereby no fact of the matter about correct translation. Blackburn’s and Searle’s claim that indeterminacy does not arise at home leads the same *kind* of mistake, i.e., blurring the crucial distinction between indeterminacy and underdetermination. It is to force Quine, who is a non-factualist about fine-grained meanings, to submit to a factualist view of such meanings.

The third problem with Searle’s and Blackburn’s reading is that they treat the speaker’s *own* language in such a way that it looks as if meanings are fixed within it and as if there is no question of translation in the speaker’s own language. They talk about the speaker’s knowing what she means by the terms of her own language and that translation does not happen at home, as if we are allowed to talk about determinate meanings without talking about correct translation. This cannot be held in Quine’s naturalistic philosophy. For Quine, “there is no entity without identity; [thus] no meaning without sameness of meaning” (1995a, 75-76). We cannot claim that something exists if we fail to provide a criterion for its individuation. We cannot talk about physical objects, for instance, if we cannot give a criterion for their identity. According to Quine, we do talk about physical

objects because we have such a criterion for them. As Kemp puts it, “Quine’s notion of a *physical object* ... has a sharp criterion of identity: x and y are the same physical object if and only if their spatio-temporal boundaries coincide” (2006, 136). What is such a criterion when the subject matter is that of meaning? We can say that a sentence has a uniquely determinate meaning only if we can tell when two sentences can share the same meaning: A and B have the same meaning if and only if they are the correct translations of one another. But Quine’s arguments for the indeterminacy of translation rejects the possibility of sustaining any such criterion of identity for the case of meaning: there is no fact as to whether two sentences are the correct translations of each other. Consequently, we cannot talk about meanings either. Therefore, we cannot use the notion of meaning independently of that of correct translation or sameness of meaning. The claims from Blackburn and Searle are thereby implausible in Quine’s view, i.e., that *in the case of the speaker herself speaking her own language*, there is no shadow of translation and indeterminacy: whenever we talk about meaning of terms, no matter if they belong to an alien language or a familiar one, we are talking about translation. And there is no fact about correct translation. Nor is any about fine-grained meanings. Therefore, the solutions that Searle and Blackburn offer are to be viewed as implausible. Hylton, however, attempts to offer a different solution.

4. Hylton’s Defense of Quine

Hylton suggests a middle view, but as we can immediately see, he seems to commit to the same *kind* of mistake that Blackburn and Searle made, because his first claim is that “[w]hile we are simply using our language, there is no issue of indeterminacy, for that arises only with translation” (2007, 214). Hylton’s claim, as it stands, is already susceptible to the three problems that were discussed above. But Hylton adds that “Quine certainly need not deny that I often know what I am about to do” (2007, 214). For him, if indeterminacy arises in the case of the translator’s own language too, the translator’s utterances would remain indeterminate, in which case we face the following vicious regress: we are trapped in endless attempts to find another language the meaning of whose expressions is supposedly determinate so that we can translate our own words into that language. Obviously, each time the indeterminacy survives. For Hylton, the regress must be stopped somewhere because it does lead to a denial of first-person authority and the best place to stop it is in our own language. But the fundamental question is: How?

The main problem here is that, as previously discussed, there seems to be no naturalistically promising answer to such a question in Quine’s philosophy. Hylton

attempts to offer an answer to this question by appealing to the following remark from Quine: “*in practice* we end the regress of background languages, in discussions of reference, by acquiescing in our mother tongue and taking its words at face value” (1969, 49, my emphases). There is an essential role that the proviso “in practice” plays in this passage, which I think is overlooked by Hylton here. What Quine seems to emphasize here is not that the regress ends by assuming that there *are* determinate meanings in our *own* language, the meaning that we, as first-persons, know. This cannot be Quine’s view. His point is rather that when there is no fact as to what the correct translation of an expression is, we are left with nothing but *pragmatic* or *practical* criteria to decide between them. Theoretically speaking, this whole idea is independent of whether we are translating an unknown language or the language of our fellow-speakers. Hylton is right in his claim that “acquiescing in our mother tongue and taking its words at face value” would, for Quine, mean that we simply *use* our own language. But this by no means helps argue that, for Quine, in the case of our own language, “translations, determinate or indeterminate, are not to the point” (Hylton 1990/91, 280). They are always to the point but not in practice. *In practice*, of course, this is true. As Quine has always insisted on, we do mean things in real life and understand what others mean by their words. But these are two separate claims. Quine has been very careful in distinguishing between the issues with indeterminacy of translation and the issues with successful linguistic communication: “the indeterminacy of translation holds also at home. But adherence to the home language can nevertheless afford escape from problems of translation” (1973, 83-84). This may be viewed as an endorsement of what Hylton, and even Blackburn and Searle, have been suggesting, i.e., that the indeterminacy problem is somehow solved at home, so that meanings are determined and ready to be known by the speakers. But Quine then adds: “It all depends on what we are trying to do” (1973, 84). And this is crucial.

Indeterminacy is ubiquitous: linguistic expressions have no determinate (fine-grained) meaning. We come up with this conclusion when “we are going to make capital of relations of sameness and difference of meaning” (Quine 1973, 84). In practice, normal speakers are not concerned with such theoretical and sceptical issues. Quine then continues: I may be equally interested in “the steps by which you all may plausibly have arrived at much the same usage as mine; but the *sameness* that I have in mind here is merely the sameness that is tested by *smoothness of dialogue*, and *not a sameness of hidden meanings*” (1973, 84, emphases added). Practically speaking, we choose among rival translations the one that helps our communication proceed more smoothly. We are so acquainted with

our home language that we speak and understand unhesitatingly and almost automatically. This, however, has nothing to do with the fact that translation and its accompanied indeterminacy hold at home. To think that meaning is somehow *determined* in our own language is to make a mistake very similar to that which Chomsky has made, at least according to Quine. If meaning and reference is, as a matter of fact and not for some pragmatic reasons, viewed as determined at home, then meaning and reference would be at most *underdetermined* rather than *indeterminate*. The whole point of Quine's remarks on the indeterminacy thesis was that, in the case of meaning and translation, we have indeterminacy *in addition to* underdetermination. There is then *no* stop to the regress which Hylton mentioned: the indeterminacy is there to assure us that it would not stop. There is no essential difference between a translator translating an unknown language and a translator communicating with, translating, or understanding the speech of her fellow speakers. If we think that things are different at home, it is because of our familiarity with the use of our own language.

All this, however, does not imply that our initial concern with Quine's view has vanished: Quine must either deny first-person authority, in which case his view remains problematic – at least for being counterintuitive – or provide some constructive remarks on how to accommodate the meaning-asymmetry. I think he can offer such a story.

5. A Wittgenstein-Strawsonian Explanation

Strawson famously distinguishes between two types of predicates, M-predicates and P-predicates. "The first kind of predicate consists of those which are also properly applied to material bodies to which we would not dream of applying predicates ascribing states of consciousness" (Strawson 1959, 104), such as "weighs 10 kg," "is 30°C," and the like; "[t]he second kind consists of all the other predicates we apply to persons" (Strawson 1959, 104), such as "believes that" and "means that."¹⁵ For Strawson, however, there should be some way for determining whether a speaker has possessed such P-predicates. What is such a criterion?

¹⁵ The concept of a person is taken by Strawson to be a primitive concept, i.e., "the concept of a type of entity such that *both* predicates ascribing states of consciousness *and* predicates ascribing corporeal characteristics, a physical situation &c. are equally applicable to a single individual of that single type" (1959, 101-102). See also Strawson (1959, 110-111). We need not be worried about the primitiveness of this concept as what we are concerned with here is Strawson's criteria for the application of P-predicates. The primitiveness of the concept of a person can lead to interesting questions, even with regard to Quine's philosophy. This would be the subject of an independent investigation. For a discussion of this, see e.g., Hacker (2002) and Snowdon (2009).

According to Strawson, “[t]here is no sense in the idea of ascribing states of consciousness to oneself, or at all, unless the ascriber already knows how to ascribe at least some states of consciousness to others” (1959, 106). The important point here is to see how an asymmetry of the sort Quine needs can emerge. On the one hand, “one ascribes P-predicates to others on the strength of observation of their behaviour” (Strawson 1959, 106). On the other, “when one ascribes them [P-predicates] *to oneself*, one does not do so on the strength of observation of those behaviour criteria on the strength of which one ascribes them to others” (1959, 107). This difference provides distinct criteria for the application of P-predicates, which in turn enable us to ascribe conscious states to ourselves *differently* from the way we ascribe them to others. This is a claim that Quine can in principle embrace: “it is essential to the character of these predicates that they have both first- and third-person ascriptive uses, that they are both self-ascribable otherwise than on the basis of observation of the behaviour of the subject of them, and other-ascribable on the basis of behaviour criteria” (Strawson 1959, 108). This explanation, if one prefers to call it so, does not need to presuppose the notion of fine-grained meanings; rather, it is enough that there is an (intersubjective) agreement in use about these terms so their use can be treated as already fixed. As Strawson famously puts it,

[t]o learn their use is to learn both aspects of their use. In order to *have* this type of concept, one must be both a self-ascriber and an other-ascriber of such predicates, and must see every other as a self-ascriber. In order to *understand* this type of concept, one must acknowledge that there is a kind of predicate which is unambiguously and adequately ascribable *both* on the basis of observation of the subject of the predicate *and* not on this basis, i.e. independently of observation of the subject: the second case is the case where the ascriber is also the subject. (1959, 108)

Through learning our first language, we gradually gain mastery of the use of our language. Our languages contain different expressions, with different rules governing their application. Such a mastery, among other things, consists in our ability to use the same type of predicates on the basis of two different criteria. Quine’s Wittgensteinian idea that language is essentially public, “a social art” (Quine 1960, x), together with his naturalized epistemology, which brings in the role of *evidence* in our descriptions of the world and persons, can prepare us to treat certain of the speaker’s utterances as self-ascriptions of meaning and as authoritative and reliable by default. Competent speakers of a language are disposed to apply “means that,” “intends that” and other P-predicates to others on the basis of evidence and observation and to themselves differently, i.e., free from any need to observe their own verbal behaviour. Those who do not possess such a

mastery cannot be said to have learnt their language: enough evidence to the contrary leads to an expiration of their membership in the speech-community.¹⁶

The account can work within Quine's naturalistic philosophy because it is *neutral* to the issues about the metaphysics of meaning, what they are and what constitutes facts about them. The desired asymmetry emerges in the following way: we ascribe meanings, *whatever they might be*, to others on the basis of evidence and observation; we do *not* do the same in our own cases. And doing so is part of what constitutes one's mastery of one's own language. This is basically what Wittgenstein pictures when he says "[w]hat is the criterion for the redness of an image? For me, when it is someone else's image: what he says or does. For myself, when it is my image: nothing" (1953, §377). What sustains such a fact is *not* some hidden fact about the fine-grained meaning of my word.¹⁷ It is rather a *primitive* aspect of being a member of a speech-community that self-ascriptions are viewed as authoritative by default. It is a form of life with its own various language-games in which we attribute meanings to ourselves and to others in different ways.¹⁸

Whether or not this sketched Strawsonian-Wittgensteinian account is viewed as plausible, to some important extent, depends on the sort of metaphysical view from which you are attempting to deal with this phenomenon. It is not a plausible account if you wish to stand outside of our best theory, or as Wittgenstein puts it, our own form of life, and take a factualist view of fine-grained meanings. If you believe there *are* such semantic facts, the account would lose its force. Davidson, for instance, has rejected this view on a similar ground.¹⁹

¹⁶ For certain well-known discussions of this issue see, e.g., Kripke (1982, Chapter 3), Dummett (1994, 423-425) and Wright (2001, 202-203).

¹⁷ Kripke's Wittgenstein has argued powerfully against the plausibility of any such attempt. See Kripke (1982, Chapter 2).

¹⁸ Whether this Wittgensteinian idea leads to what Strawson calls the "no-ownership" view (1959, 95) is a matter of controversy. See, e.g., footnote 1 in Strawson (1959, 95). Strawson later emphasizes that "[w]hat I am suggesting is that it is easier to understand how we can see each other, and ourselves, as persons, if we think first of the fact that we act, and act on each other, and act in accordance with a common human nature" (1959, 112). Comparing to Wittgenstein's "Beetle in a Box" (1953, §293), for instance, it does not seem that there appears a serious conflict between these two views. For Strawson, the word "I" does not refer to some Cartesian sort of pure conscious self, but "refers to the *person* who uses it" (Hacker 2002, 25). This, I believe, leads to no conflict with Wittgenstein's view, especially considering his remarks on the primitiveness of forms of life. Nonetheless, deciding on this controversy would be subject to an independent investigation. For Strawson's reading of Wittgenstein's relevant remarks, see Strawson (1974/2008, Especially Chapter 7). For more on Strawson's view, see Bilgrami (2006, Chapter 2; 2010), Baldwin (2010), Davidson (1984; 1993b), Hacker (2002), and Snowdon (2009).

¹⁹ See, e.g., Davidson (1984; 1987; 1993b).

His reasons may seem convincing but only if we, following him, dramatically depart from Quine's naturalistic philosophy. He looks at the problem from a factualist point of view.²⁰ For Davidson, "we may postulate different criteria of application for the key concepts or words ('believes that,' 'intends to,' 'wishes that,' etc.). But these moves do no more than restate the problem" (1984, 109). He thinks that the asymmetry in ascriptions of meaning cannot be explained by an appeal to the existence of different criteria of use. For, first of all,

it is a strange idea that claims made without evidential or observational support should be favored over claims with such support. Of course, if evidence is not cited in support of a claim, the claim cannot be impugned by questioning the truth or relevance of the evidence. But these points hardly suffice to suggest that in general claims without evidential support are more trustworthy than those with. (1987, 442)

But is it a strange idea? For one thing, Strawson does *not* think that "claims made without evidential or observational support should be favored over claims with such support;" he rather, not strangely at all, claims that *mastery* of how to use our language would, among other things, consist in knowing how to use its predicates, that some of its predicates are used without an appeal to evidence and some inevitably by an appeal to such evidence. Furthermore, what Strawson, and I think Quine, needs would be the platitude that "if evidence is not cited in support of a claim, the claim cannot be impugned by questioning the truth or relevance of the evidence." The speaker's claim is that "I mean *rabbit* by 'rabbit'" and she knows, among other things, that when she uses "mean that" in self-ascriptions, she uses it on the basis of a criterion different from the one governing the use of it in the case of ascribing meaning to others. It is a strange idea for Davidson because he thinks we have a better way to explain the asymmetry in question: we can appeal to the fact that understanding, or more particularly, interpreting speakers in the way they intended their utterances to be understood, was not be possible if they did not have non-inferential knowledge of what they mean and believe.²¹ Davidson views facts about meaning, beliefs and intentions as real as any other commonly conceded facts: he *includes such facts in the class of all facts about the world*. As he puts it, propositional attitudes are "as real as atoms and baseball bats, and the facts about them are as real as the facts about anything else" (1998, 98). This is the reason why

²⁰ On his departure from Quine on this matter, see Hossein Khani (2018; 2023), Kemp (2012, 127), and Engel (2013, 594).

²¹ On Davidson's account, see, e.g., Aune (2012), Child (2007), Hacker (1997), Hossein Khani (2021b), and Ludwig and Lepore (2005, 353-354).

he thinks that Quine's indeterminacy thesis "does not entail that there are no facts of the matter" (1999, 596).

Davidson's more sophisticated objection is that Strawson's account results in an ambiguity in the meaning of P-predicates and this would invite scepticism about other minds. According to Davidson, "if what is apparently the same expression is sometimes correctly employed on the basis of a certain range of evidential support and sometimes on the basis of another range of evidential support (or none), the obvious conclusion would seem to be that the expression is ambiguous" (1987, 442). In that case, he asks: "Why then should we suppose that a predicate like 'x believes that...', which is applied sometimes on the basis of behavioral evidence and sometimes not, is unambiguous?" (1987, 442). And if it is ambiguous, it does not seem that the predicate preserves its meaning when employed by the speaker in self-ascriptions of attitudes and by the speaker in the case of ascribing them to others. This leads to scepticism about other minds because the sceptic is then allowed to ask whether the self-ascribed attitudes really are the same as those we attributed to others. On this basis, Davidson objects that "Strawson and Wittgenstein had *described* the asymmetry, but had done nothing to explain it" (1993a, 211).²² Nonetheless, the existence of differences in use is part of what a mastery of the use of language is, in which case it would then be a "strange idea" to claim that such a use is *ambiguous*, in such a Davidsonian sense, rather than simply *rule-governed*. And rules are not supposed to be simple: they can be as complex as Strawson thinks and "unambiguously" applicable if we are to be able to judge if one has learnt one's language.

Quine would not be against any of these claims. Such an intersubjective ground, agreement across a speech-community, has already been present in Quine's definition of observation sentences – as those sentences that almost all members of the speech-community have more or less similar dispositions to assent or dissent on specific occasions.²³ The sketched account here appeals only to what Quine strongly admits, i.e., that "knowing what expressions mean *consists*, for me, in being disposed to use them on appropriate occasions" (Quine 2000, 420). And the differences in such dispositions would enable Quine to preserve the asymmetry in question. I disagree with Hylton's claim that, for Quine, "meaningfulness *consists in* just the sorts of things that make translation possible" (1990/91, 274). If Quine wanted to follow this path of reasoning, he should have said *what makes*

²² See also Davidson (1993b, 248-249).

²³ His later discussion of pre-established harmony is relevant but in a complicated way. See Quine (1995b; 1996). See also Kemp (2017).

translation possible at all. But for Quine, nothing philosophically salient can be said about such a possibility.²⁴

References

- Aune, Bruce. 2012. "On Davidson's View of First-Person Authority." In *Donald Davidson on Truth, Meaning, and the Mental*, edited by G. Preyer, 214-227. Oxford: Oxford University Press.
- Baldwin, Thomas. 2010. "Comments on A.K. Bilgrami's Self-Knowledge and Resentment." *Philosophy and Phenomenological Research* LXXXI (3): 773-782.
- Bilgrami, Akeel. 2006. *Self-Knowledge and Resentment*. Cambridge, MA: Harvard University Press.
- Bilgrami, Akeel. 2010. "Precis of Self-Knowledge and Resentment." *Philosophy and Phenomenological Research* LXXXI (3): 749-765.
- Blackburn, Simon. 1984. *Spreading the Word*. Oxford: Oxford University Press.
- Child, William. 2007. "Davidson on First Person Authority and Knowledge of Meaning." *Noûs* 41 (2): 157-177.
- Chomsky, Noam. 1968. "Quine's Empirical Assumptions." *Synthese* 19 (1/2): 53-68.
- Davidson, Donald. 1984. "First Person Authority." *Dialectica* 38: 101-112.
- Davidson, Donald. 1987. "Knowing One's Own Mind." *Proceedings and Addresses of the American Philosophical Association*, 60 (3): 441-458.
- Davidson, Donald. 1993a. "Reply to Eva Picardi." In *Reflecting Davidson*, edited by R. Stoecker, 210-212. Berlin: Walter de Gruyter & Co.
- Davidson, Donald. 1993b. "Reply to Bernard Thöle." In *Reflecting Davidson*, edited by R. Stoecker, 248-250. Berlin: Walter de Gruyter & Co.
- Davidson, Donald. 1998. "Replies." *Critica* 30: 97-112.
- Davidson, Donald. 1999. "Reply to Richard Rorty." In *The Philosophy of Donald Davidson*, edited by L. E. Hahn, 595-600. Illinois: Open Court.
- Dummett, Michael. 1994. "Testimony and Memory." In *Knowing from Words*, edited by B. K. Matilal and A. Chakrabarti, 251-272. Amsterdam: Kluwer.
- Engel, Pascal. 2013. "Davidson and Contemporary Philosophy." In *A Companion to Donald Davidson*, edited by Ernie Lepore and Kirk Ludwig, 588-604. Malden, MA.: Wiley-Blackwell.

²⁴ **Disclosure statement:** The author reports there are no competing interests to declare. **Conflict of Interest:** The author declares no competing interests.

- Føllesdal, Dagfinn. 1990. "Indeterminacy and Mental States." In *Perspectives on Quine*, edited by R. B. Barrett and R. F. Gibson, 98-109. Cambridge, MA: Basil Blackwell.
- Føllesdal, Dagfinn. 2014. "Developments in Quine's Behaviorism." In *A Companion to W.V.O. Quine*, edited by G. Harman and E. Lepore, 263-278. West Sussex, UK: Wiley & Sons Ltd.
- Gibson, Roger. 1986. *The Philosophy of W.V. Quine: An Expository Essay*. Florida: University Press of Florida.
- Gibson, Roger. 2004. *The Cambridge Companion to Quine*. Cambridge: Cambridge University Press.
- Glock, Hans-Johann. 2003. *Quine and Davidson on Language, Thought, and Reality*. Cambridge: Cambridge University Press.
- Hacker, Peter. 1997. "Davidson on First-Person Authority." *Philosophical Quarterly* 47: 285-304.
- Hacker, Peter. 2002. "Strawson's Concept of a Person." *Proceedings of the Aristotelian Society* 102: 21-40.
- Harman, Gilbert. 2014. "Indeterminacy, Relativity, and Behaviorism." In *A Companion to W.V.O. Quine*, edited by G. Harman and E. Lepore, 219-235. West Sussex, UK: Wiley & Sons Ltd.
- Hookway, Christopher. 1988. *Quine*. Cambridge: Polity Press.
- Hosseini Khani, Ali. 2018. "Quine on the Indeterminacy of Translation: A Dilemma for Davidson." *Dialectica* 72 (1): 101-120.
- Hosseini Khani, Ali. 2021a. "The Indeterminacy of Translation and Radical Interpretation." *The Internet Encyclopedia of Philosophy (IEP)*. <https://iep.utm.edu/indeterm>
- Hosseini Khani, Ali. 2021b. "Davidson on Self-Knowledge: A Transcendental Explanation." *The Southern Journal of Philosophy* 59 (2): 153-184.
- Hosseini Khani, Ali. 2023. "The Root of the Third Dogma of Empiricism: Davidson vs. Quine on Factualism." *Acta Analytica* 38 (1): 161-183.
- Hylton, Peter. 1982. "Analyticity and the Indeterminacy of Translation." *Synthese* 52: 167-184.
- Hylton, Peter. 1990/91. "Translation, Meaning, and Self-Knowledge." *Proceedings of the Aristotelian Society* 91: 269-290.
- Hylton, Peter. 2007. *Quine*. NY: Routledge.
- Kemp, Gary. 2006. *Quine: A Guide for the Perplexed*. London: Continuum.
- Kemp, Gary. 2012. *Quine versus Davidson: Truth, Reference, and Meaning*. Oxford: Oxford University Press.

- Kemp, Gary. 2017. "Quine, Publicity, and Pre-Established Harmony." *ProtoSociology* 34: 59-72.
- Kirk, Robert. 1969. "Translation and Indeterminacy." *Mind* 78 (311): 321-341.
- Kirk, Robert. 1986. *Translation Determined*. Oxford: Oxford University Press.
- Kripke, Saul. 1982. *Wittgenstein on Rules and Private Language*. Cambridge, MA: Harvard University Press.
- Ludwig, Kirk, and Ernie Lepore. 2005. *Donald Davidson: Meaning, Truth, Language, and Reality*. Oxford: Oxford University Press.
- McGinn, Marie. 1997. *The Routledge Guidebook to Wittgenstein's Philosophical Investigations*. NY: Routledge.
- Orenstein, Alex. 2002. *W. V. Quine*. Acumen.
- Quine, W.V. 1948. "On What There Is." *The Review of Metaphysics* 2 (1): 21-38.
- Quine, W.V. 1960. *Word and Object*. Cambridge, MA: MIT Press.
- Quine, W.V. 1968. "Reply to Chomsky." *Synthese* 19 (1/2): 274-283.
- Quine, W.V. 1969. *Ontological Relativity and Other Essays*. New York: Columbia University Press.
- Quine, W.V. 1970. "On the Reasons for Indeterminacy of Translation." *The Journal of Philosophy* 67 (6): 178-183.
- Quine, W.V. 1973. *The Roots of Reference*. Open Court.
- Quine, W.V. 1975. "On Empirically Equivalent Systems of the World." *Erkenntnis* 9: 313-328.
- Quine, W.V. 1978. "Facts of the Matter." *The Southwestern Journal of Philosophy* 9 (2): 155-169.
- Quine, W.V. 1981. *Theories and Things*. Cambridge, MA: Harvard University Press.
- Quine, W.V. 1986. "Reply to John Woods." In *The Philosophy of W. V. Quine*, edited by E. H. Hahn and P. A. Schilpp, 726-728. Chicago: Open Court.
- Quine, W.V. 1987. "Indeterminacy of Translation Again." *The Journal of Philosophy* 84 (1): 5-10.
- Quine, W.V. 1990a. *Pursuit of Truth*. Cambridge, MA: Harvard University Press.
- Quine, W.V. 1990b. "Comment on Parsons." In *Perspectives on Quine*, edited by Roger Gibson and R. B. Barrett, 291-293. Oxford: Blackwell.
- Quine, W.V. 1990c. "Three Indeterminacies." In *Perspectives on Quine*, edited by Roger Gibson and R. B. Barrett, 1-16. Blackwell.
- Quine, W.V. 1995a. *From Stimulus to Science*. Cambridge, MA.: Harvard University Press.

- Quine, W.V. 1995b. "Preestablished Harmony." In *Quine and His Place in History (2016)*, edited by Janssen-Lauret F. and Kemp G., 29-36. London: Palgrave Macmillan.
- Quine, W.V. 1996. "Progress on Two Fronts." *The Journal of Philosophy* 93 (4): 159-163.
- Quine, W.V. 2000. "Quine's Responses." In *Knowledge, Language and Logic*, edited by Alex Orenstein and Petr Kotatko, 407-430. Dordrecht: Kluwer Academic.
- Searle, John. 1987. "Indeterminacy, Empiricism, and the First Person." *The Journal of Philosophy* 84 (3): 123-146.
- Snowdon, Paul. 2009. "Persons' and Persons." *Organon F* 16 (4): 449 – 476.
- Soames, Scott. 1999. "The Indeterminacy of Translation and the Inscrutability of Reference." *Canadian Journal of Philosophy* 29 (3): 321-370.
- Strawson, Peter. 1959. *Individuals*. London: Methuen.
- Strawson, Peter. 1974/2008. *Freedom and Resentment and Other Essays*. New York: Routledge.
- Verhaegh, Sander. 2018. *Working from Within: The Nature and Development of Quine's Naturalism*. Oxford: Oxford University Press.
- Wittgenstein, Ludwig. 1953. *Philosophical Investigations*. Translated by G.E.M. Anscombe. Oxford: Blackwell.
- Wittgenstein, Ludwig. 1958. *The Blue and Brown Books*. Oxford: Wiley Blackwell.
- Wright, Crispin. 1992. *Truth and objectivity*. Cambridge, MA: Harvard University Press.
- Wright, Crispin. 2001. *Rails to Infinity: Essays on Themes from Wittgenstein's Philosophical Investigations*. Cambridge, MA: Harvard University Press.

WHY THE HECK WOULD YOU DO PHILOSOPHY? A PRACTICAL CHALLENGE TO PHILOSOPHIZING

Jimmy Alfonso LICON

ABSTRACT: Philosophy plausibly aims at knowledge; it would thus be tempting to hold that much of the value of doing philosophy turns on securing knowledge. Enter the agnostic challenge: suppose that a philosophical agnostic (named ‘Betsy’) wants to discover only fundamental philosophical truths. However, the intractable disagreement among philosophical experts gives her pause. After reflecting on expert disagreement, she decides that doing philosophy, *for her truth-seeking error-avoiding purposes*, is irrational. In this paper, I argue that the agnostic challenge isn’t easily overcome. Although there are many reasons to do philosophy, the agnostic challenge implies there is less value to doing philosophy than many philosophers may have believed.

KEYWORDS: agnostic challenge, the intractable disagreement,
philosophical experts

[Philosophers] are not honest enough in their work, although they make a lot of virtuous noise when the problem of truthfulness is touched even remotely. They all pose as if they had discovered and reached their real opinions through the self-development of a cold, pure, divinely unconcerned dialectic [...]; while at bottom it is an assumption, a hunch, indeed a kind of “inspiration” [...]—that they defend with reasons they have sought after the fact.

—Friedrich Nietzsche, *Beyond Good and Evil*

1. Introduction¹

There is a serious *prima facie* challenge to the practical value of doing philosophy: intractable disagreement among philosophical experts² exists over answers to

¹ Thanks to Julius Schoenherr, Jake Sheen, and a referee for helpful feedback on prior drafts.

² Someone might worry that this appraisal of disagreement among philosophers is too strong; it might not be that *all* parties, or *even most parties*, in the dispute are informed about the arguments relevant to the dispute. However, it is sufficient for my purposes that there are *some*

fundamental philosophical questions despite the fact that ‘... arguments thought relevant to the disputed questions are... [often] well-known to all parties to the dispute’ (Kelly 2005, 173). Such experts hold numerous, mutually exclusive answers to fundamental philosophical questions. It is thus reasonable that truth-seeking, error-avoiding agnostics should be disturbed by intractable disagreement.

The phenomenon of intractable disagreement among philosophers is striking, especially since many philosophers hold that ‘philosophy, like all other studies, aims primarily at knowledge’ (Russell 2001 [1912], 90). Wilfred Sellars once wrote that the aim of philosophy ‘is to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term’ (2007, 369). Even many contemporary philosophers implicitly take the value of doing philosophy to be, in part, a truth-finding, error-avoiding enterprise. It is difficult to suppose that if such philosophers took seriously the prospect that doing philosophy would be an unreliable, or perhaps dubious approach, to securing fundamental philosophical truths that they would continue to see the value of doing philosophy as intact.

Enter the philosophical agnostic: after weighing this omnipresent and apparently intractable disagreement, we may imagine that our agnostic (call her ‘Betsy’) would conclude that she isn’t better situated epistemically, to find answers to such questions, than philosophical experts. And she concludes that she should remain agnostic about philosophical matters. The philosophical literature backs up Betsy’s impulse: many skeptical challenges to the epistemic credentials of philosophical claims have been well defended by philosophers (Goldberg 2009, 2013; Licon 2012, 2013; Kornblith 2013; Frances 2016).

Barriers to the epistemic credentials of philosophical claims, along with plausible means-ends reasoning, implies that the agnostic³ has good reason to conclude that doing philosophy *for the purposes of seeking truth and avoiding falsehood to an equal degree* is irrational (i.e., it would be a waste of time). By implication, doing philosophy *ceteris paribus* is less valuable than we might have otherwise supposed, if one holds that there value of doing philosophy is, at least to a large degree, a product of discovering philosophical truths. If philosophizing isn’t a viable means of finding philosophical truths, the value of doing philosophy is partly undermined.⁴ Call this *the agnostic challenge*. We can illustrate the agnostic challenge with the following story:

parties in the dispute who are familiar enough with the arguments on all sides of the dispute who still disagree with each other.

³ The focus on agnostics should not be confused with a focus on philosophical *novices*.

⁴ These issues relate to whether philosophy makes progress (Stoljar 2017; Chalmers 2015).

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

Betsy is a sincere philosophical agnostic who wants to find truths, and avoid falsehoods, about fundamental philosophical issues. However, Betsy recognizes that philosophical experts hold a many incompatible answers to such questions. Betsy concludes that such intractable dissensus is good evidence that she would be no better epistemically situated to find true answers herself. But if so, doing philosophy to find truths and avoid falsehoods about fundamental philosophical issues would be seemingly intractable. Betsy concludes that this seemingly intractability makes doing philosophy, *for her purposes*, practically irrational (e.g., it would be a waste of time).

In this paper, I argue that there are no obviously satisfactory answers to the agnostic challenge. I take Betsy as my paradigm agnostic. In a nutshell, the agnostic challenge traffics in skepticism about the value of doing philosophy to find fundamental truths, and avoid falsehoods, about philosophical questions—the agnostic challenge doesn’t target the value of doing philosophy *simpliciter*. There are a number of reasons doing philosophy may be valuable (e.g., philosophical inquiry might be intrinsically valuable) that are compatible with the agnostic challenge. There may be philosophical experts who find that reasonable disagreement with their colleagues is sufficiently valuable such that doing philosophy is worthwhile. Even still, it would be worrying if we cannot convince a reasonable agnostic like Betsy that doing philosophy *for the purposes of finding truth and avoiding falsehood* is a worthwhile pursuit. Such a failure should significantly diminish the value of doing philosophy.

2. The Agnostic Challenge

The agnostic challenge has two facets: worries that settling philosophical disputes is intractable that tie into the prudential worry that attempting to solve an intractable philosophical dispute, for the sake of trying to solve it, is practically irrational. And in section (A), I briefly defend the practical claim. Then in sections (B) and (C), I defend the claim that philosophizing in order to find philosophical truths and avoid error is likely an intractable task, especially with respect to fundamental⁵ philosophical issues. Finally, I address possible answers to the agnostic challenge and find them (mostly) wanting.

A. IRRATIONALITY AND INTRACTABLE PROBLEMS

At the heart of the agnostic challenge is the following prudential claim:

⁵ Agnostics like Betsy might accept that doing philosophy allows one to discover certain kinds of philosophical truths—e.g., justification cannot transform true belief into knowledge—but this concession wouldn’t satisfy the agnostic challenge.

WASTE If F is an intractable problem, trying to solve F, *for the sake of solving* F, is irrational (Bratman 1998, 61–7; Finkelstein 2007).

By way of motivating WASTE, we need to briefly think about the irrationality of trying to solve intractable problems: it is irrational to attempt solving a problem we take to be intractable solely for the sake of trying to solve it. The point is not that trying to solve an intractable problem *for enjoyment* is irrational, otherwise playing Tetris would be pointless. Instead, it is trying to solve an intractable problem that one has good reason to believe is intractable, solely for the sake of trying to solve the problem, is practically irrational.

There are eccentric individuals with different prudential values who will fail to feel the pull of WASTE. Peter enjoys feeling frustrated by the process of problem solving; he is never satisfied by the prospect of solving a problem, not to mention *actually* solving it. In order to enhance his frustration, and thus his enjoyment, he inhibits his ability to solve problems. Peter enjoys trying to solve a problem without making progress toward solving it⁶. Even if one fails to find WASTE plausible, Betsy is rational to accept it—WASTE might not be rationally compelling, but it is rationally acceptable. It seems reasonable to hold that trying to solve an apparently intractable problem, solely for the sake of solving the apparently intractable problem, is irrational.

B. FUNDAMENTAL PHILOSOPHICAL QUESTIONS

We may wonder how to characterize a fundamental philosophical issue. Philosophers entertain lots of questions, many of which look piecemeal (e.g., how best to characterize Fregean senses). It may be unclear how to distinguish between *fundamental* and *non-fundamental* philosophical questions. However, it is not vital to precisely capture the nature of fundamental philosophical questions for a couple of reasons—though we should be able to *roughly* characterize them as a rough characterization is what is minimally needed to motivate the agnostic challenge. First, we can characterize them *operationally*; and second, much of what philosophy itself purports to do involves answering questions fundamental to the concerns of agnostics like Betsy, e.g., how we could have knowledge of abstract objects? Along with Chalmers (2015), I frame fundamental philosophical questions as,

[Questions] of philosophy [...] like: What is the relationship between mind and body? How do we know about the external world? What are the fundamental

⁶ If one supposes that such cases aren't possible (i.e. trying to phi, while trying not to succeed at phi-ing, is not a coherent possibility), then all the better for the agnostic challenge.

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

principles of morality? Is there a god? Do we have free will? [...] any philosopher can come up with a list of 10 or so big questions fairly easily, and I suspect that there would be a lot of overlap between these lists. We could even use these lists to operationally define the big questions: the big questions of a field at time t are those that members of that field would count as the big questions of the field at time t (5).

Someone might worry that if we characterize fundamental problems operationally, then what counts as a fundamental problem might change from generation of philosophers to the next; in such a short time, it might be unreasonable to expect that finding true answers to fundamental philosophical questions would be intractable. While one would be right to press this point, it will only take us so far: many questions in contemporary philosophers would no doubt count as *fundamental* bear a striking similarity to many questions that have occupied many philosophers since Plato and Aristotle (even if many questions *don't* possess this resemblance). For example, questions about the evil of death, the nature of right and wrong, the nature of knowledge, the relationship between mind and body, etc., are questions philosophers still strive to answer. It is fundamental philosophical questions of broadly this sort that interest agnostics like Betsy.

C. INTRACTABLE DISAGREEMENT AMONG PHILOSOPHERS

Why think finding truths about fundamental philosophical issues is intractable? One answer to this question involves appealing to apparently intractable philosophical disagreement. Though disagreement *per se* isn't epistemically worrisome—otherwise, we should be skeptical about nearly everything we believe, as there are undoubtedly folks somewhere who would disagree—certain kinds of disagreement are epistemically worrisome: mutually inconsistent positions held by philosophical experts is reason to worry about the reliability of philosophical methods.

If we assume philosophical experts are (roughly) equally competent using such methods, then their intractable disagreement is reason to doubt the reliability of such methods. If the methods were used correctly, and they were reliable, then we should expect plenty more convergence among philosophers than we find. Either some experts are inept with regard to philosophical methods—such a possibility doesn't help since Betsy doesn't have good reason to suppose she would be more competent than experts—or philosophical methods are unreliable. Either way then, agnostics like Betsy have a serious *prima facie* reason not to suppose that doing philosophy is not a reliable means of arriving at the truth. However, the epistemic situation gets worse: we need not hold that philosophical methods are

unreliable to produce a defeater for philosophical beliefs. Even if philosophical methods are reliable, we would still have a reason to suppose that they aren't reliable in cases of intractable disagreement, and those cases are omnipresent among philosophers (Goldberg 2009, 109-11; Kornblith 2013; Barnett 2017).

Agnostics like Betsy lack good reason to think that they would be better epistemically situated to find truths and avoid falsehoods about fundamental philosophical issues than philosophical experts. A truth-seeking, error-avoiding philosophical *novice* would have no reason to invest the time and energy required to become a philosophical expert if she'll be no better epistemically situated to find answers than the current experts. She would likely end up disagreeing with her colleagues, but without a means to fruitfully resolve the disagreement.

Although there isn't nearly as much empirical evidence, regarding the extent of disagreement among philosophical experts, as we might like, there is enough to strongly suggest that experts are deeply divided: out of the thirty questions that contemporary philosophers answered, there was at least *weak* convergence (roughly sixty percent) on seven out of the thirty questions (e.g., normative ethics is almost perfectly divided between the consequentialists, deontologists, virtue ethicists, and others: Bourget and Chalmers 2014). Not every question on the survey cited are *fundamental* philosophical questions; however, it is striking how little agreement there is among philosophers, compared to the sciences or math, on questions that are central to the discipline, especially in light of the lifetime of effort philosophers that devote to these questions (Goldberg 2013; Schwitzgebel and Cushman 2015, 2012; Earlenbaugh and Molyneux 2009).

Worse still, to the limited extent that there is consensus in philosophy, we might worry that this is sometimes due to epistemically irrelevant factors. For instance, Ballantyne (2014) and Licon (2013) argue that the fact that there are philosophers in nearby possible worlds who disagree with us may pose a skeptical challenge to the justificatory standing of our philosophical beliefs. They are, roughly, cases of disagreement between actual and merely modal philosophers, in which the mere fact that such philosophers aren't actual is epistemically irrelevant, and this ties into the literature on irrelevant influences on belief (Vavova 2018; White 2010). The subject is too developed to properly survey here; we need only appreciate that even given convergence, agnostics like Betsy may still grapple with the agnostic challenge for counterfactual reasons—disagreement is epistemically worrying, but agreement isn't always reassuring.

Even if experts hold reasonable, but incompatible positions (Feldman 2007, Decker 2012) such a prospect doesn't answer the agnostic challenge. Agnostics like Betsy aren't doing philosophy simply to find reasonable answers to fundamental

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

philosophical questions; they could accept that there are many instances of reasonable disagreement, without supposing that such cases settle the agnostic challenge. Betsy might even grant that experts reasonably disagree with each other—the worry here, though, is that reasonable belief doesn't preclude false belief, and thus doesn't address the (equally important for agnostics) issue of avoiding falsehoods.

3. Answering the Challenge

In this section, we'll examine a few answers to the agnostic challenge. However, as I will argue, such answers are unsatisfactory either individually or together in that they should fail to satisfy Betsy's primary concerns. While such answers sometimes capture the value of philosophizing, e.g., the intrinsic value of weighing philosophical reasons, their failure to answer the agnostic challenge should worry professional philosophers and agnostics alike. If philosophizing cannot uncover truths to fundamental questions, then doing philosophy is less valuable than many philosophers would have otherwise supposed.

A. THE SELF-DEFEAT ANSWER

There is an obvious worry that the agnostic challenge is self-defeating. Although the agnostic challenge isn't a fundamental philosophical question, it relies on philosophical methods which appear, given apparently intractable disagreement, to have questionable epistemic credentials. If Betsy has reason to worry that philosophical methods are epistemically suspect, we should worry that the philosophical methods used to motivate the agnostic challenge are epistemically suspect, thereby casting doubt on the agnostic challenge itself. However, despite the initial look of plausibility, there are a few problems with this answer to the agnostic challenge.

First, suppose for the sake argument that the agnostic challenge is right: we have good reason to worry that philosophical methods are poorly equipped to deliver fundamental philosophical truths. An implication would be that one couldn't motivate the agnostic challenge as that would be self-defeating. However, if philosophy cannot muster the resources to motivate the agnostic challenge, then we should find that as troubling as the failure to answer the agnostic challenge to begin with—if philosophical methods are so unreliable, they can't even motivate the agnostic, then the epistemic credentials of those methods are in worse epistemic shape than we originally supposed. This suggests that the self-defeat answer faces the following dilemma:

Either philosophical methods are enough to motivate the agnostic challenge, even

if they cannot satisfactorily answer that challenge; Or, such methods aren't sufficient to motivate the agnostic challenge—but then, Betsy should remain an agnostic with regard to philosophical issues, since philosophical methods are too facile to mount a challenge to philosophical practices.

On the first horn of the dilemma, philosophical methods are sufficient to produce reasonable⁷ beliefs, and that in turn is sufficient to get the agnostic challenge off the ground. Agnostics like Betsy aren't worried about reasonable philosophical beliefs. She is worried that it is the task of attempting to find truths, and avoid falsehoods, about fundamental philosophical questions is irrational. On the first horn of the dilemma, agnostics like Betsy can use philosophical methods to pose a viable challenge to the practical value of doing philosophy by appealing to reasonable philosophical beliefs. On the other horn: philosophical methods cannot even justify reasonable philosophical beliefs. And this should be more worrying than the agnostic challenge itself since it appears that philosophical methods are even feebler than originally supposed.

Second, the self-defeat answer misapprehends the agnostic challenge: the challenge doesn't take a stance on whether philosophical methods can yield reasonable philosophical beliefs (Feldman 2007; Decker 2012). The agnostic challenge is about *truth and falsehood*. Agnostics like Betsy do not claim that the agnostic challenge is true, but instead that the agnostic challenge is *reasonable*. Betsy and her agnostic ilk can consistently hold that the agnostic challenge is *reasonable* while holding it offers reason to avoid do philosophy. The challenge neither denies the value of doing philosophy simpliciter, nor that one can hold reasonable philosophical beliefs given informed disagreement. Betsy holding that the agnostic challenge is philosophically reasonable isn't self-defeating⁸ (especially given meta-philosophical disagreement: Kornblith 2013).

B. THE BIRTHING ANSWER

Philosophy has (allegedly) birthed scientific disciplines instrumental in unearthing answers to questions that were once thought philosophical in nature. If so, then doing philosophy might allow philosophers to indirectly find true answers to questions formerly thought philosophical. On this answer, only philosophical

⁷ Philosophical methods may provide *justified* philosophical beliefs; but Betsy worries that there is too much space between justified and truths about fundamental philosophical issues for that to be a satisfying answer to the agnostic challenge (Lehrer and Cohen 1983).

⁸ The apparent self-defeat may not by itself be reason to reject an argument (Matheson 2015; Lammenranta 2011). Those who want to defend the self-defeat answer must say more about the alleged self-defeating nature of the agnostic challenge, even if we grant it is self-defeating.

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

questions that remain intractable would be the purview of philosophy, while tractable questions would be taken over by science (Chalmers 2015, 25-6). If agnostics like Betsy want to uncover fundamental philosophical truths, then they should do philosophy to birth novel that indirectly reveal where to find fundamental philosophical truths. And while there is passing plausibility to this answer, there are a few reasons that Betsy should be dissatisfied with it.

First, the birthing answer is as contentious as answers to fundamental philosophical questions. Although philosophers often pay lip service to the idea that philosophy has given birth to new disciplines throughout its history—e.g., physics and economics—they rarely say why we should accept this claim⁹. It isn't that Betsy must *know* that the birthing answer is right, but if she wants to avoid falsehoods and find truths about fundamental philosophical questions equally, then she would be susceptible to false answers to fundamental philosophical questions, and also to wasting her time and energy, if the birthing answer is false. And worse still, many fundamental philosophical questions are unanswered despite philosophy birthing disciplines like economics, psychology, and physics. Agnostics like Betsy shouldn't hold their breath, even if the birthing answer is broadly correct.

Second, even if the birthing answer is broadly right, it only tells us that it is *possible* that doing philosophy could birth a new discipline that could indirectly answer fundamental philosophical questions. But if the odds of an agnostic like Betsy birthing a new discipline that would help find true answers to such questions were five million-to-one, where she would only succeed once out of five million attempts, she would *still* have good reason to suppose that doing philosophy is irrational: she seriously risks wasting her time without anything to show for it. Even if doing philosophy to birth new disciplines would be fruitful in answering such questions is *possible*, this insufficient to hold that doing philosophy is a good bet for birthing disciplines—even if philosophy births new disciplines, it only happens rarely.

Third, the birthing answer must assume that fundamental philosophical questions are disguised *non-philosophical* questions—philosophers devised them, and scientists answer them. If either (a) some fundamental philosophical questions must be answered using philosophical methods, or (b) some questions are not amenable to philosophical methods, even if they are philosophical in nature, then at best, agnostics Betsy should be partially satisfied. The birthing answer would only give her reason to hold that some fundamental philosophical questions will be

⁹ The paucity of literature on this question should itself be troubling (especially given how often philosophers pay lip service to the claim).

answered once there are new, better equipped disciplines to answer such questions have. And, putting all that aside, it still wouldn't be a sufficient reason *to do philosophy* as a means to find truths about fundamental philosophical questions.

Finally, even where established sciences interact with philosophical questions, this is at most, a partial interaction (Chalmers 2015, 10-11): the sciences may constrain answers that philosophers may plausibly defend but without resolving the philosophical aspects of the issue (Mele 2014; Carruthers 2009). Philosophical aspects of fundamental questions often stay intractable, despite advances in the salient sciences. Though the sciences inform and constrain plausible answers to fundamental philosophical questions, such questions frequently retain philosophical aspects that remain susceptible to intractable disagreement among philosophers.

Here a critic may object that whether the birthing answer is correct is likely an empirical issue, and thus isn't a fundamental philosophical question. And if it is an empirical matter, agnostics like Betsy would be better off waiting for the empirical evidence to settle the issue, instead of treating the lack of resolution as reason to stay philosophically agnostic. However, the problem with this objection is that the birthing answer isn't simply resolvable empirically, and likely has philosophical and conceptual issues intertwined with the empirical side of the issue such that Betsy may still find herself facing intractable disagreement with philosophical colleagues.

C. THE SPECTRUM ANSWER

This Quinean viewpoint is that philosophy is on the same spectrum as the sciences, except more abstract and general than the sciences (Quine 1981, 67). However, if that is correct, the agnostic challenge wrongly focuses on one part of the intellectual spectrum that includes philosophy, while discounting the enormous success of the science part of the spectrum, then concludes that given intractable disagreement on the *philosophical part of the spectrum*, it would be irrational for agnostics like Betsy to do philosophy. In a nutshell, it appears that the agnostic challenge might require arbitrarily dividing up the academic spectrum, then pointing to the many failures by philosophers to find truths about fundamental philosophical issues in fashion that is arbitrary—if philosophy is on the same spectrum as science, then shouldn't the successes of science count (at least indirectly) for something with regard to philosophy? And despite the initial plausibility of this answer, it should be less than convincing to agnostics like Betsy for several reasons.

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

First, even if philosophy is on the same intellectual spectrum as science, this is beside the point with regard to the agnostic challenge. To see why, consider that we could rephrase the agnostic challenge: from the perspective of agnostics like Betsy, segments of the intellectual spectrum do better (the science end of the spectrum) than others (the philosophy end of the spectrum) with regard to finding truths to fundamental questions such disciplines investigate¹⁰. So, if anything, the spectrum answer gives one a reason to do science if they want to make progress on certain kinds of questions. But that doesn't give Betsy reason to do *philosophy* to find the truths about fundamental questions in philosophy. Though it may be that the scientific end of the intellectual spectrum would be a place to make progress in answering certain kinds of questions, it is not clear how this would help agnostics like Betsy.

Second, we should doubt the assumption that underlies the spectrum answer: many methods used by philosophers are *prima facie* distinct from methods used by scientists. This isn't to doubt that there could be some methodological overlap: e.g., philosophers and scientists use thought experiments. Despite some similarities though, philosophical and scientific methods are distinct in many areas. Philosophical tools include intuitions, counterexamples, reflective equilibrium and so on, while scientific tools include qualitative, statistical, and so on. This doesn't deny the distinction between disciplines is a bit fuzzy; however, it is reason to deny that belonging on the same spectrum means that philosophy should get credit for the successes of science.

Third, the spectrum answer is nearly as controversial as answers to fundamental philosophical questions. Suppose that Betsy adopted the spectrum answer: she still would have no idea which part of the spectrum to focus on. While there is seemingly a great deal of progress on the hard sciences end of the spectrum, physics doesn't appear concerned with fundamental *philosophical* questions. Perhaps this is because philosophical questions are disguised scientific questions, for instance. From Betsy's epistemic vantage point, though, physics doesn't appear better placed to help her uncover truths to such questions than accounting (with perhaps a few exceptions). On the other hand, if Betsy adopts philosophical methods to such questions, she will engage with questions that concern her, but at the expense of finding answers she can have confidence are true. Betsy can either get results to questions that do not appear philosophy, or she can work on philosophical questions with good reason to doubt that she would

¹⁰ Even this claim is dubious: many philosophical questions that are not *prima facie* amenable to scientific inquiry. Of course, this might be a false impression, but it is a plausible one.

find truths about such issues. Either avenue would be equally unsatisfying to agnostics like Betsy.

D. THE DESTRUCTION ANSWER

Perhaps philosophical methods uncover truths about fundamental questions indirectly via the process of elimination. On this view, philosophical progress would occur by elimination: once we identify wrong views, then we can know by elimination which answers are true based on whatever answers remain. For example, if we have solid reason to believe that God existence is impossible,¹¹ then by the process of elimination we have good reason to adopt atheism. If we could do philosophy in a fashion that would reliably eliminate falsehoods, where only truths remain, Betsy would have good reason to pursue philosophy: it would uncover truths about fundamental philosophical questions via evolutionary pressures. And while there is something plausible to this answer, there are a couple reasons Betsy shouldn't buy it.

First, philosophy doesn't viably shorten the list of viable answers to fundamental philosophical questions to a single answer, even though that would seem to be a prerequisite to satisfying the agnostic challenge. If philosophical methods reliably eliminated all but several viable answers, then philosophers would have little to disagree about. We often reject positions in philosophy, but we rarely get a single viable answer. So, Betsy would have to settle for plausible answers to such questions. However, this is insufficient to answer the agnostic challenge.

Second, even if philosophy shortened the list of viable answers to fundamental philosophical questions to a single answer, we should worry that this destruction only makes room for more sophisticated versions of previously discarded positions. For example, the moral skeptic doesn't convince the moral realist to disavow her metaethical realism, but instead motivates the moral realist to devise a more sophisticated position. As Lycan (2013) notes:

Philosophical consensus is far more the result of Zeitgeist, fad, fashion, and careerism than of accumulation of probative argument [...] as a game, I was once challenged by Red Watson to exhibit the arguments that refuted sense-datum theory. I spent an hour or two marshalling them. He swiftly and decisively pointed out that each of them either begs the question or at least has a premise that would not impress an actual sense-datum theorist in the slightest. And *not because the sense-datum theorist was being dogmatic* (116-7, emphasis mine).

¹¹ Some philosophers have made this argument, but with little success: Martin and Monnier (2003).

Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing

The same could be claimed of most debates in philosophy: they pressure interlocutors to devise more sophisticated versions of their position to respond to criticism. However, this doesn't help agnostics like Betsy. It is inadequate to merely discredit answers to fundamental philosophical questions unless the product is a single viable answer, while precluding the possibility of more sophisticated competing answers arising as a result. For example, many philosophers hold that there are definitive objections to substance dualism; however, other philosophers, in response, have devised more sophisticated defenses of the view (Lycan 2009; Molyneux 2015).

4. Conclusion

The agnostic challenge is a serious challenge to the practical value of doing philosophy: if one is a truth seeking, error-avoiding agnostic, then doing philosophy for the sake of finding answers to fundamental philosophical questions is irrational. Agnostics like Betsy have little reason to hold that they would be more likely to find truths and avoid falsehoods about fundamental philosophical issues, than philosophical experts. The agnostic challenge is weighty reason for philosophers to worry that doing philosophy isn't as valuable as they may have otherwise believed, especially since many philosophers hold that philosophical inquiry is valuable as an avenue to truth. If doing philosophy isn't a viable means of finding fundamental philosophical truths, then much of the value of doing philosophy is undermined.

References

- Ballantyne, N. 2014. "Counterfactual Philosophers." *Philosophy and Phenomenological Research* 88 (2): 368-387.
- Barnett, Z. 2017. "Philosophy without Belief." *Mind* 128 (509): 109-138 <doi.org/10.1093/mind/fzw076>
- Bergmann, M. 1997. "Internalism, Externalism and the No-Defeater Condition." *Synthese* 110 (3): 399-417.
- Bourget, D. & D.J. Chalmers. 2014. "What do Philosophers Believe?" *Philosophical Studies* 170 (3): 465-500.
- Bratman, M. E. 1998. "Toxin, Temptation, and the Stability of Intention." In *Rational Commitment and Social Justice*, eds. J. L. Coleman and C. W. Morris, 59-83. Cambridge University Press.
- Carruthers, P. 2009. "How We Know Our Own Minds: The Relationship Between Mindreading and Metacognition." *Behavioral and Brain Sciences* 32 (2): 121-138.

Jimmy Alfonso Licon

- Chalmers, D. J. 2015. "Why Isn't There More Progress in Philosophy?" *Philosophy* 90 (1): 3-31.
- Decker, J. 2012. "Disagreement, Evidence, and Agnosticism." *Synthese* 187 (2): 753-783.
- Earlenbaugh, J. & B. Molyneux. 2009. "Intuitions are Inclinations to Believe." *Philosophical Studies* 145 (1): 89-109.
- Feldman, R. 2007. "Reasonable Religious Disagreements." In *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, ed. Louise Antony, 194-214. Oxford University Press.
- Finkelstein, C. 2007. "Rational Temptation." In *Practical Rationality and Preference*, eds. Christopher W. Morris and Arthur Ripstein, 56-80. Cambridge University Press.
- Frances, B. 2016. "Worrisome Skepticism about Philosophy." *Episteme* 13 (3): 289-303.
- Goldberg, S. 2009. "Reliabilism in Philosophy." *Philosophical Studies* 142 (1): 105-117.
- . 2013. "Defending Philosophy in the Face of Systematic Disagreement." In *Disagreement and Skepticism*, ed. Diego E. Machuca, 277-294. Routledge.
- Kelly, T. 2005. "The Epistemic Significance of Disagreement." In *Oxford Studies in Epistemology*, Vol. 1, eds. John Hawthorne and Tamar Gendler, 167-196. Oxford University Press.
- Kornblith, H. 2013. "Is Philosophical Knowledge Possible?" In *Disagreement and Skepticism*, ed. Diego E. Machuca, 260-276. Routledge.
- Lammenranta, M. 2011. "Disagreement, Skepticism, and the Dialectical Conception of Justification." *International Journal for the Study of Skepticism* 1 (1): 3-17.
- Lehrer, K. & S. Cohen. 1983. "Justification, Truth, and Coherence." *Synthese* 55 (2): 191-207.
- Licon, J.A. 2012. "Sceptical Thoughts on Philosophical Expertise." *Logos and Episteme* 3 (3): 449—458.
- . 2013. "On Merely Modal Epistemic Peers: Challenging the Equal-Weight View." *Philosophia* 41 (3): 809-823.
- Lycan, W. G. 2009. "Giving Dualism Its Due." *Australasian Journal of Philosophy* 87 (4): 551-563.
- . 2013. "On Two Main Themes in Gutting's What Philosophers Know." *Southern Journal of Philosophy* 51 (1): 112-120.
- Martin, M. & R. Monnier. 2003. *The Impossibility of God*. Prometheus Books.

- Why the Heck Would You Do Philosophy? A Practical Challenge to Philosophizing
- Matheson, J. 2015. "Are Conciliatory Views of Disagreement Self-Defeating?" *Social Epistemology* 29 (2): 145-159.
- Mele, A. R. 2014. *Free: Why Science Hasn't Disproved Free Will*. Oxford University Press USA.
- Molyneux, B. 2015. "Zombies and Probability: A Stochastic Approach to the Dualist's Problems." *American Philosophical Quarterly* 52 (3): 201-213.
- Quine, W. V. (ed.). 1981. *Theories and Things*. Harvard University Press.
- Russell, B. 2001 [1912]. *The Problems of Philosophy*. Oxford University Press.
- Schwitzgebel, E. & F. Cushman. 2012. "Expertise in Moral Reasoning? Order Effects on Moral Judgment in Professional Philosophers and Non-Philosophers." *Mind and Language* 27 (2): 135-153.
- . 2015. "Philosophers' Biased Judgments Persist Despite Training, Expertise and Reflection." *Cognition* (141): 127-137.
- Stoljar, D. 2017. *Philosophical Progress: In Defence of a Reasonable Optimism*. Oxford University Press.
- Vavova, K. 2018. "Irrelevant Influences." *Philosophy and Phenomenological Research* 96 (1):134-152.
- White, R. 2010. "You just believe that because..." *Philosophical Perspectives* 24 (1): 573-615.

OUGHT WE TO BELIEVE THE TRUTH AND NOTHING BUT THE TRUTH? TWO ARGUMENTS FOR THE WIDE SCOPE VERSION OF THE TRUTH NORM

Daniel RÖNNEDAL

ABSTRACT: According to the so-called truth norm, we ought to believe that A if and only if A is true. There are many possible interpretations of this norm. What does 'ought' in this norm mean? Does 'ought' have a wide or a narrow scope, etc.? In this paper, I will investigate one version of this norm and I will discuss two arguments for it. The 'ought' in the paper will be interpreted as a kind of 'rational' ought that takes wide scope. I will call the first argument for the truth norm 'the extrapolation argument' and the second argument 'the abductive argument.' According to the extrapolation argument, we 'derive' the truth norm from a reflection on what it means to be a perfect believer. According to the abductive argument, the truth norm is supported by the fact that it can be used to deduce many other plausible doxastic norms. If this argument is successful, the truth norm can be conceived as *the* fundamental norm of (theoretical) rationality (or wisdom).

KEYWORDS: the truth norm, the aim of belief, the consistency norm, rational requirements, rationality, wisdom

1. Introduction

Ought we to believe the truth and nothing but the truth? And what does this proposition mean? In recent years, such questions have often been discussed in connection with another idea, namely the idea that the aim of belief is the truth.¹ Some have defended this thesis in one form or another,² some have criticised it.³

¹ See, for example, Atkinson (2021), Boghossian (2003), Bykvist and Hattiangadi (2007), Chan (2013), Côté-Bouchard (2016), Gibbard (2003, 2005), McHugh (2011, 2012), McHugh and Whiting (2014), Owens (2003), Raleigh (2013), Shah (2003), Shah and Velleman (2005), Steglich-Petersen (2006, 2009, 2013), Toribio (2013), Vahid (2006), Velleman (2000), Wedgwood (2002, 2007), Whiting (2010, 2012, 2013), Yamada (2012), Zalabardo (2010), and Zangwill (2005).

² See, for example, Boghossian (2003), Engel (2007, 2013), Fassio (2011), Shah and Velleman (2005), Wedgwood (2002, 2007) and Whiting (2010).

³ See, for example, Bykvist and Hattiangadi (2007, 2013), Glüer and Wikforss (2013), and Papineau (2013).

But what does it mean to say that the aim of belief is the truth? Obviously, this is a metaphor that can be interpreted in many ways (Fassio 2022). According to several philosophers (see, for example, the introduction in Chan 2013), the idea that the truth is the aim of belief means that the following proposition is valid:

(CN). The correctness norm. It is correct to believe that A if and only if (iff) it is true that A.

A number of thinkers have argued that ‘correctness’ is a deontic term that can be defined in terms of what ought to be the case (in some sense) (see, for example, Boghossian 2003). Hence, many philosophers also accept some version of the so-called truth norm, which tells us to believe the truth and nothing but the truth:

(TN). The truth norm. We ought to believe that A iff A is true.

However, some seem to believe that the correctness norm is primitive or more fundamental than the truth norm (see, for example, Wedgwood 2002, 2007, 2013). Some have argued for a truth norm defined in terms of what *may* be the case. According to them, it is permitted that we believe that A iff A is true (see, for example, Whiting 2010). Some believe that the principle should be expressed in evaluative rather than deontic terms (see, for example, McHugh 2012).

The truth norm can be divided into two parts:

(TN1) We ought to believe the truth, and

(TN2) We ought to believe nothing but the truth.

Obviously, the conjunction of (TN1) and (TN2) is equivalent with (TN). Some are inclined to accept (TN2) (in some sense) but not (TN1) (see, for example, Boghossian 2003).

There are many possible interpretations of the truth norm. For example, does ‘ought’ have a wide or a narrow scope? Should we read it as ‘(We ought to believe that A) iff A is true’ or as ‘It ought to be that (we believe that A iff A is true)’? What kind of ‘ought’ is involved in the truth norm? Is it a moral ought, a prudential ought or some other kind of ought? Is it some kind of ‘rational,’ ‘doxastic’ or ‘epistemic’ ought?

Several philosophers have criticised the truth norm (see, for example, Bykvist and Hattiangadi 2007). They have, for example, argued that this principle violates the ‘ought’ implies ‘can’ principle and that it is inconsistent with the truth of some so-called ‘blindspot propositions.’ Such arguments seem to refute some versions of the truth norm, in particular some narrow scope versions, but it is not obvious that they can be used to show that every reading of this norm is false.

Ought We to Believe the Truth and Nothing But the Truth?

In this paper, I will interpret the truth norm as a wide scope norm that involves a kind of ‘rational,’ ‘doxastic’ or ‘epistemic’ ought. This version seems to avoid many problems with the narrow scope interpretation, and it is possible to find some interesting arguments for it (see Sections 2 and 3). Here is our rendition of the principle:

(WTN). The wide scope truth norm. (Insofar as rationality is concerned) it ought to be that (we believe that A iff A [is true]).

This norm can be divided into two parts:

(WTNI). (Insofar as rationality is concerned) it ought to be that (if A [is true], then we believe that A).

(WTNII). (Insofar as rationality is concerned) it ought to be that (we believe that A only if A [is true]).

Instead of ‘Insofar as rationality is concerned, it ought to be that A’ we can use the expressions ‘Rationality requires that A’ or ‘Rationally, it ought to be the case that A.’ However, to keep things simple, I will usually only talk about what ‘ought to be the case,’ and when I do, I usually mean what ‘ought to be the case insofar as rationality is concerned.’ The word ‘ought’ can be used in many different senses; rational, prudential and moral requirements are logically independent. So, it can, for example, be true that rationality requires that A even though morality requires that not-A.⁴

Here is an example of an instance of (WTNI): It ought to be that (if $2 + 2 = 4$, then we believe that $2 + 2 = 4$), and here is an example of an instance of (WTNII): It ought to be that (we believe that $2 + 2 = 4$ only if $2 + 2 = 4$).

I will discuss two arguments for (WTN) in this paper. In Section 2, I will introduce an argument that I will call ‘the extrapolation argument,’ and in Section 3, I will investigate an argument that I will call ‘the abductive argument.’ According to the extrapolation argument, we ‘derive’ the truth norm from a reflection on what it means to be a perfect believer. And according to the abductive argument, the truth norm is supported by the fact that it can be conceived as the fundamental norm of (theoretical) rationality (or wisdom). Section 4 is a short conclusion.

⁴ Readers who think that rationality only has to do with consistency or means-end efficiency or have a narrow concept of rationality might try rephrasing (WTN). Instead of speaking about what ‘rationality requires’ we can perhaps talk about what the doxastic norms or wisdom or the aim of belief requires. The exact words are not that important. The important thing is that there seems to be a special kind of doxastic or epistemic norms that are different from other kinds of norms, for example, moral norms. (WTN) is supposed to be a norm of this kind.

2. The Extrapolation Argument

It is often valuable to believe what is true and not to believe what is false. Suppose that it is true that there is a lion in the bush. You do not believe that this is the case. So, you stroll along without any fear. The lion attacks you and kills you. Obviously, in this case it would have been good to believe the truth. Suppose that you believe that it is safe to drink the water from this pond (perhaps because it has been safe before). Since you believe this and are thirsty, you drink from the pond. However, your belief is false. The water is poisonous. As a consequence, you die a painful death.

These examples suggest that there are two things we want: to believe what is true and to avoid believing what is false. If we do not believe what is true (for example, that there is a lion in the bush) or if we believe something that is false (for example, that it is safe to drink the poisonous water in the pond), all sorts of negative consequences might follow. Now, we extrapolate from this. What would it mean to be a perfect believer or a perfectly rational or wise person? Well, it is natural to think that a perfect believer or a perfectly rational or wise person would believe everything that is true and that she would believe nothing that is false. A perfectly wise person, in this sense, would not accidentally get eaten by lions or drink from poisonous ponds.

We form the following hypotheses:

(W1). Wise 1. It is necessary that for every individual S, S is perfectly (theoretically) wise or rational (a perfect believer) only if, for every proposition P, if P is true then S believes P.

(W2). Wise 2. It is necessary that for every individual S, S is perfectly (theoretically) wise or rational (a perfect believer) only if, for every proposition P, S believes that P only if P is true.

Suppose now that everyone ought to be perfectly (theoretically) wise or rational (a perfect believer). Then we can derive (WTN) from (W1) and (W2). The only thing we have to assume is that rational requirements transfer over necessary implications. That is, the only thing we must assume is that if it is necessary that A implies B, then it ought to be the case that B if it ought to be the case that A. And this seems to be eminently plausible. I believe that something like this might be one (although not the only) route to the truth norm.

So, have we proven that the truth norm is 'valid'? Even though the extrapolation argument seems quite interesting to me, there are some potential problems with it. I will mention two.

According to the first problem, the argument is based on an overgeneralization. It is often valuable to believe what is true and not to believe

what is false. But this is not always the case. Sometimes it seems to be bad (or neither good nor bad) for us to believe what is true and good (or neither good nor bad) for us not to believe what is true. Suppose that there are 1.324.784 blades of grass in this part of the lawn. Is it good for everyone to believe this? Is it bad if not everyone believes this proposition? Is it bad if anyone (falsely) believes that there are 1.324.783 blades of grass in this part of the lawn? Suppose S's partner has cheated on S. Is it good for S to believe that her partner has been unfaithful? Is it bad if S (falsely) believes that her partner is faithful? Is it bad if S does not believe that her partner has cheated on her? So, maybe a perfect believer would not believe the truth and nothing but the truth.

A defender of the argument might point out that it is difficult to draw the line between truths (falsehoods) that a perfect believer will believe and truths (falsehoods) that a perfect believer will not believe. If a perfect believer does not believe every truth, which truths does she believe? When can a perfect believer believe something false? Any way of drawing the line seems arbitrary. There can be prudential, moral etc. reasons not to believe the truth and nothing but the truth. But maybe a perfect *believer* will satisfy (W1) and (W2).

According to the second problem, 'ought' implies 'can.' If 'ought' implies 'can,' we ought to be perfectly (theoretically) wise or rational (a perfect believer) only if we can be perfectly (theoretically) wise or rational (a perfect believer). Therefore, we ought to be perfectly (theoretically) wise or rational (a perfect believer) only if we can be perfectly (theoretically) wise or rational (a perfect believer). But it is not the case that we can be perfectly (theoretically) wise or rational (a perfect believer). Consequently, it is not the case that we ought to be perfectly (theoretically) wise or rational (a perfect believer). Hence, a crucial premise in the extrapolation argument is false. It follows that the extrapolation argument fails.

If perfect (theoretical) wisdom or rationality entails that a perfectly wise or rational individual believes that A iff A is true, then the premise that we cannot be perfectly (theoretically) wise or rational is very plausible. The only reasonable way to avoid this counterargument seems to be to reject the 'ought' implies 'can' principle for rational requirements (for a similar response to a counterargument to the abductive argument, see Section 3 below).

A possible response to the counterarguments above is to restrict (W1), (W2) and (WTN). In Section 3, we will mention some possible versions of (WTN). However, such a defence of the extrapolation argument seems much less plausible than a similar defence of the abductive argument. So, even though the

extrapolation argument is quite interesting it is not a watertight proof of the truth norm.

3. The Abductive Argument

Several philosophers have suggested or hinted at the idea that the truth norm is *the* fundamental epistemic or doxastic norm (see, for example, Boghossian 2003; Engel 2013; Wedgwood 2002, 2007). However, it is not always clear what they mean by this, and the idea has not been developed in detail. According to one interpretation of this thought, it means that other epistemic or doxastic norms in some sense depend on the truth norm.

If the truth norm is indeed a fundamental norm, how can one argue for it? If it is a basic principle, it seems that it should be treated like an axiom, and there are no deductively valid arguments for (genuine) axioms. So, we should not expect to be able to find any deductively valid arguments for the truth norm. Instead, we must look at the consequences of the norm and how it coheres with other beliefs. Does it have reasonable consequences? Does it have any problematic implications?

Nevertheless, we can formulate an abductive argument for the truth norm. An abductive argument has the following form: B. If A were the case, then B would be the case. Hence A. Obviously, such an argument is not deductively valid. The idea is rather that B in some sense supports A, or that we have good (fallible) reason to believe A if our abductive argument is strong.

It is possible to derive many different 'epistemic' or 'doxastic' norms from the truth norm, for example the following principle: we ought not to combine believing A and believing that A implies B with not believing B. However, I will focus on just *one* example in this paper. I will show how the so-called consistency norm follows from the truth norm. This is a widely accepted norm that is intuitively plausible. It has, for example, been explicitly defended by Harry Gensler (see Gensler 1986, 1996, Chapter 2). The consistency norm includes two parts (just as the truth norm). According to the first part, it ought to be that you do not combine inconsistent beliefs. More precisely, the first part can be formulated in the following way:

(WCNI). The (wide scope) consistency norm (for two beliefs) (Part I). If A is inconsistent with B, it ought to be that you do not combine believing A with believing B.

Here is an example of an instance of Part I: If the proposition that all swans are white is inconsistent with the proposition that this swan is not white, then it ought to be the case that you do not combine believing that all swans are white with believing that this swan is not white.

This norm can be generalized in the following way. If a set of propositions is inconsistent, then you ought not to believe every proposition in this set. Or more precisely, if $\{A_1, \dots, A_n\}$ (for $n \geq 1$) is inconsistent, then you ought not to combine believing A_1, \dots , and believing A_n . This principle can be called ‘The (wide scope) consistency norm (for n beliefs) (Part I)’. However, to keep things simple, I will focus on the consistency norm.

Why shouldn’t we have inconsistent beliefs? Here is one possible answer. If it is impossible that A and B and you believe A and you believe B , then there is at least one proposition that you believe that is false. If you believe something that is false, you violate Part II of the truth norm, that is, the principle that we ought to believe nothing but the truth. So, the truth norm can explain why we shouldn’t have inconsistent beliefs.

We can show that the consistency norm (Part I) follows from the truth norm (Part II) in a rigorous way. We only have to assume that our modal and deontic expressions function as normal modal and deontic operators. For more on modal and deontic logic, see, for example, Garson (2018), and McNamara, and Van De Putte (2021). Here is our proof.

Suppose that (WTNII) is true and that (WCNI) is not true in some possible world w . Then it is impossible that A and B in w , and it is not the case that it ought to be that you do not believe that A and you believe that B in w . Since it is not the case that it ought to be that you do not believe that A and you believe that B in w , there is a possible world w' that is deontically accessible from w in which you believe that A and you believe that B . Since (WTNII) is true in w it is true in w' that it ought to be that you believe something only if it is true. Since w' is deontically accessible from w it follows that ‘if you believe that A then A ’ is true in w' and that ‘if you believe that B then B ’ is true in w' . Hence, A is true in w' and B is true in w' . Therefore, A and B is true in w' . But A and B is not true in w' , for A and B is impossible in w . This is absurd. Hence, it is not possible that (WTNII) is true and that (WCNI) is not true in some possible world w . Consequently, (WCNI) follows from (WTNII).

According to the second part of the consistency norm, it ought to be that you do not believe something without believing what is necessarily implied by it. More precisely, the second part can be expressed in the following way:

(WCNII). The (wide scope) consistency norm (for two beliefs) (Part II). If A necessarily implies B , it ought to be that you do not combine believing A with not believing B .

Here is an example of an instance of Part II: If the proposition that all swans are white necessarily implies that this swan is white, then it ought to be that you do

not combine believing that all swans are white with not believing that this swan is white.

Part II of this norm can also be called 'the consequence norm' or 'the closure norm.' It can be generalized in the following way. If a set of premises necessarily implies a conclusion, then you ought not to believe every proposition in this set without believing the conclusion. Or more precisely, if $\{A_1, \dots, A_n\}$ (for $n \geq 1$) necessarily implies B, then you ought not to combine believing A_1, \dots , and believing A_n with not believing B. This principle can be called 'The (wide scope) consistency norm (for n beliefs) (Part II).' Nevertheless, to keep things simple, I will focus on the consistency norm.

Why should you believe what is necessarily implied by your beliefs? If it is necessary that A implies B and you believe A and you do not believe B, then there is at least one proposition that you believe that is false or else there is at least one true proposition that you do not believe. So, if you violate (WCNII), either you violate Part I or Part II of the truth norm, i.e. the norm that we should believe what is true or the norm that we should believe only what is true.

The consistency norm (Part II) follows from the truth norm (Part I and Part II). This can be shown in a rigorous way. Here is our proof.

Suppose that (WTNI) and (WTNII) are true and that (WCNII) is not true in some possible world w. Then it is true in w that it ought to be that if A then you believe that A and it is true in w that it ought to be that you believe A only if A. Since (WCNII) is false in w it is true that it is necessary that A implies B in w and it is false that it ought to be that you do not combine believing A with not believing B in w. Since it is not true that it ought to be that you do not combine believing A with not believing B in w, there is a possible world w' that is deontically accessible from w in which it is true that you believe A and it is false that you believe B. Since (WTNII) is true in w it is true that if you believe A then A in w'. And since (WTNI) is true in w it is true in w' that if B then you believe B. Hence, A is true in w'. Since it is necessary that A implies B in w it is true that A implies B in w'. Consequently, B is true in w'. It follows that it is true in w' that you believe B. But this is absurd. Hence, it is not possible that (WTNI) and (WTNII) are true and that (WCNII) is not true in some possible world w. Consequently, (WCNII) follows from (WTNI) and (WTNII).

(WCN). The (wide scope) consistency norm (for two beliefs) is the conjunction of (WCNI) and (WCNII).

So, the consistency norm follows from the truth norm. Since we have good reason to believe that the consistency norm is true, it follows that we have good reason to believe in the truth norm.

The basic intuition behind the abductive argument can also be explained in the following way. Why should we care about consistency? Answer: Because we should care about the truth. Consistency is a necessary condition for believing the truth and nothing but the truth (this can be shown rigorously; the proof is left to the reader). Gensler (1986), for example, treats the consistency principle as an axiom. But if the abductive argument is sound, the consistency principle can be explained in terms of an even more fundamental principle, namely the truth norm.

So, does this prove that the truth norm is 'valid'? Since it is always possible to reject the conclusion in any deduction if we reject at least one premise, it is not surprising that there are some potential problems with this argument. A critic could argue that the consistency norm is not true. The fact that the truth norm entails the consistency norm is therefore not any reason to accept the truth norm. I will now consider three possible reasons why some may believe that the consistency norm is not true.

According to the first reason, the consistency norm is a narrow scope norm, and not a wide scope norm. (WCN) is not a good interpretation of the consistency norm. This principle should instead be construed as the conjunction of (NCNI) and (NCNII) below:

(NCNI). The (narrow scope) consistency norm (for two beliefs) (Part I). If A and B are inconsistent, then if you believe A you ought not to believe B.

(NCNII). The (narrow scope) consistency norm (for two beliefs) (Part II). If A necessarily implies B, then if you believe A then you ought to believe B.

However, there are several serious problems with the narrow scope interpretation of the consistency norm (see Gensler 1996, Chapter 2). Let us consider some of these.

Firstly, (NCNI) can have implausible consequences. Suppose that person S believes A and also believes B and that A and B are inconsistent. Assume that (NCNI) is true. Then S ought not to believe B. For S believes A and A is inconsistent with B. Since S believes B and B is inconsistent with A, S ought not to believe A. Hence, S ought not to believe A and S ought not to believe B. But this is implausible. Even though A and B are inconsistent, it is not reasonable to conclude that S ought to give up both beliefs. Consequently, (NCNI) is not true. On the other hand, (WCNI) only tells S not to believe both A and B. If A and B are inconsistent, either A or B is false, but one of the propositions may still be true.

Secondly, (NCNI) entails that everyone that believes a contradiction (a proposition that is inconsistent with itself) ought to believe nothing. Assume that (NCNI) is true. Suppose that person S believes that A and that A is a contradiction. Then for any proposition B it is true that B is inconsistent with A. Since S believes

A and B is inconsistent with A, S ought not to believe B. Hence, for any proposition B, S ought not to believe B. That is, S ought not to believe anything. But this is absurd. Surely, it is not the case that S ought to believe nothing. Therefore, (NCNI) is not true. (WCNI) seems to be more plausible. It does not entail that everyone that believes in a contradiction ought to believe nothing. Rather, it entails that one should not believe A if A is a contradiction.

Thirdly, (NCNII) can have problematic consequences. Suppose that (NCNII) is true. Furthermore, assume that A necessarily implies B and that person S believes A and also believes not-B. Then S ought to believe B. For A entails B and S believes A. Since A necessarily implies B, not-B necessarily implies not-A. Hence, S ought to believe not-A. For S believes not-B and not-B necessarily implies not-A. Consequently, S ought to believe not-A and S also ought to believe B. That is, S ought to believe the opposite of both S's beliefs. But this is absurd. Hence, (NCNII) is not true. (WCNII) does not have any consequences of this kind. (WCNII) only tells S that S ought not to believe a proposition without believing its consequences (or necessary implications).

Fourthly, (NCNII) also entails that everyone who believes a contradiction ought to believe everything (including every contradiction). Suppose that person S believes A and that A is a contradiction. A contradiction necessarily implies everything. Hence, for any proposition B, A necessarily implies B. Since S believes A and A necessarily implies B, S ought to believe B. Hence, for any proposition B, S ought to believe B. In other words, S ought to believe everything. But this is absurd. Surely, it is not the case that S ought to believe everything. It follows that (NCNII) is not true. (WCNII) does not entail anything similar.

In all these respects, the wide scope version of the consistency norm seems to be more plausible than the narrow scope version. Therefore, this reason to reject the wide scope consistency norm does not seem to succeed.

According to the second reason, the consistency norm cannot be satisfied. It has been argued that the truth norm violates the 'ought' implies 'can' principle (see, for example, Bykvist, and Hattiangadi 2007). And since the 'ought' implies 'can' principle is true, it follows that the truth norm is not true. Bykvist and Hattiangadi express their argument against the wide scope truth norm in the following way:

take the conjunction of all the necessary truths - a proposition that is far too complex for you to grasp. According to [the truth norm], you ought to either bring it about that p is false or bring it about that you believe that p. But you can do neither. You cannot bring it about that p is false because p is a necessary truth. And you cannot come to believe that p because it is not humanly possible to grasp such a complex proposition, let alone believe it. Since 'ought' implies 'can,' [the

Ought We to Believe the Truth and Nothing But the Truth?

truth norm] must be false. (Bykvist and Hattiangadi 2007, 284–285)

A similar argument can be used against the consistency norm. The conjunction of all necessary truths is a necessary truth. A necessary truth is necessarily implied by any proposition. So, if you ought to believe anything you ought to believe the conjunction of all necessary truths, given that the consistency norm is true. But you cannot believe the conjunction of all necessary truths. ‘Ought’ implies ‘can’ and if ‘ought’ implies ‘can,’ you ought to believe the conjunction of all necessary truth only if you can believe this proposition. This is absurd, for surely you ought to believe at least something. Hence, the consistency norm is not true.

Here is another example that uses the ‘ought’ implies ‘can’ principle. It may be psychologically impossible that S believes every truth, and it may be psychologically impossible that S believes every necessary implication of what S believes. S may, for example, (truly) believe that everyone who was on the plane is dead and (truly) believe that S’s wife was on the plane. From this it follows that S’s wife is dead. S realizes that the proposition that his wife is dead follows from the proposition that everyone who was on the plane is dead and the proposition that S’s wife was on the plane. Yet, it is psychologically impossible for S to believe that his wife is dead. S is in shock. Therefore, it is not the case that S ought to believe that his wife is dead. Insofar as morality (and prudence) is concerned, it is not the case that S ought to believe that his wife is dead. Nor is it all-things-considered the case that he ought to believe that his wife is dead. And S certainly ought not to be blamed for not believing that his wife is dead. Furthermore, S cannot stop believing that everyone who was on the plane is dead, and S cannot stop believing that his wife was on the plane. Consequently, it is not the case that S ought to believe every necessary implication of what S believes. It follows that the consistency norm is not true.

How might a defender of the consistency norm respond to this counterargument? Let us consider two possible replies. According to the first response, we should restrict the consistency norm. We can, for example, modify (WCNI) and (WCNII) in the following way (a restriction of this kind was suggested already by Gensler 1986):

(RWCNI). The restricted (wide scope) consistency norm (for two beliefs) (Part I).
If A is inconsistent with B, then, insofar as you are able, you ought not to combine believing A with believing B.

(RWCNII). The restricted (wide scope) consistency norm (for two beliefs) (Part II).
If A necessarily implies B, then, insofar as you are able, you ought not to combine believing A with not believing B.

The (restricted) wide scope consistency norm, (RWCN), is the conjunction of (RWCNI) and (RWCNII). (RWCN) does not violate the 'ought' implies 'can' principle. Hence, a defender of the abductive argument and the consistency norm can avoid the 'ought' implies 'can' arguments by accepting (RWCN) instead of (WCN). A potential problem with this response is that it might be difficult to find a similarly restricted version of the truth norm that can be used to derive (RWCN). And if (WCN) should be restricted, then surely the truth norm should be restricted in a similar way.

According to the second response to the 'ought' implies 'can' argument, we should reject the 'ought' implies 'can' principle for norms of rationality. Consider our example with S and his wife. In this case, both prudence and morality permit that S does not believe that his wife is dead. But remember that the 'ought' in the truth norm and in the consistency norm is a special rational (epistemic or doxastic) 'ought.' It might seem reasonable to say that rationality requires that S believes that his wife is dead, even though it is psychologically impossible for S to believe this. Hence, rationality can require someone to believe something that she cannot believe. Therefore, the 'ought' of rationality does not imply 'can.' This idea is compatible with the plausible proposition that both prudential and moral requirements imply 'can'.

According to the third reason, the consistency norm is not true since there may be prudential and/or moral reasons not to be consistent. It would, for example, clutter our minds if we tried to satisfy the truth norm or the consistency norm. We would have to believe many things that neither interests us nor should interest us (see Harman 1986). Should we believe every necessary truth? If so, it seems that we should do nothing but work out the necessary implications of our beliefs all day long. Should we believe every 'trivial' empirical truth? Do we have to try to find out how many leaves there are on all the trees in this forest? Should we count the stars? Is it obligatory that we keep track of how many times we breathe in and out every day? Etc. Seeking information is sometimes costly, at least in time and effort. Hence, it is not always prudent.

Some truths would (potentially) be very painful for us to believe. Likewise, it would (potentially) be very painful for us to believe some consequences of what we believe. Do we have to believe every truth about every horrible crime that has ever been committed? If we have a dominant gene for an incurable disease, ought we to believe that this is the case? Suppose that x is not the biological father of y but that both x and y believe that this is the case. Should x and y revise their beliefs?

Ought We to Believe the Truth and Nothing But the Truth?

Sometimes we do not want to know the truth. We do not want to know how the book or the movie ends. We do not want to know who won (before watching the match). We take pleasure in not knowing exactly what will happen. Knowing the end in advance would, in at least some cases, spoil the experience.⁵

Perhaps we have a right to privacy. So, maybe there are truths that we should neither know nor believe. Maybe I shouldn't know what you write in your mails or your diary.

The following thought experiment suggests that the consistency norm and the truth norm are not true. Suppose that there is an extremely smart but amoral scientist who is able to scan your brain and determine whether or not you satisfy the truth norm and the consistency norm. Suppose also that this scientist is going to drop an atom bomb on a large city if you satisfy the truth norm or the consistency norm in a particular case with respect to a certain set of propositions. Assume also that it is in your power not to satisfy the truth norm and the consistency norm in this case. You only have to take a harmless drug that is provided by the scientist. If you do not take the drug, you will satisfy the truth norm and the consistency norm and the scientist will drop the bomb. If you take the drug, you will not satisfy these norms and the scientist will not harm anyone. Obviously, in this case you should not satisfy either the truth norm or the consistency norm. Consequently, neither the truth norm nor the consistency norm is true.

Examples of this kind suggest that both the consistency norm and the truth norm are false. It is not always the case that we ought to believe something iff it is true, it is not the case that we never should have inconsistent beliefs and it is not always the case that we should believe the necessary implications of what we believe.

How might a defender of the consistency norm and the truth norm respond? I will mention two possible replies. According to the first response, we should restrict the norms. This can be done in several ways. We can, for example, introduce one or two new conditions to our consistency norm of the following kind: 'if prudence permits that A' and 'if morality permits that A.' Here is an alternative formulation (a restriction of this kind was suggested already by Gensler 1986):

(CWCNI). The conditional (wide scope) consistency norm (for two beliefs) (Part I). If A is inconsistent with B, then, insofar as there are no overriding counterbalancing reasons, you ought not to combine believing A with believing B.

⁵ For more on 'pragmatic' reasons for belief, see, for example, Reisner (2009).

(CWCNII). The conditional (wide scope) consistency norm (for two beliefs) (Part II). If A necessarily implies B, then, insofar as there are no overriding counterbalancing reasons, you ought not to combine believing A with not believing B.

If we restrict the consistency norm (and similar norms), rationality cannot require something that is not prudentially permitted, and it cannot require something that is not morally permitted. Then it is not the case that we ought to believe every trivial or painful truth. It is not the case that we ought to know everything in advance. It is not the case that we should be consistent if this has catastrophic consequences etc. So, by restricting our norms, we can avoid problems of the kind discussed above.

A potential problem with this response is that it is not obvious how to restrict the truth norm in such a way that it is still possible to derive our restricted consistency norm from this principle. Consequently, it might be difficult to work out all the details of this move.

According to the second response, we should emphasise the distinction between rational requirements and prudential and moral requirements. Even though prudence or morality may require us not to be consistent, for example, we are still *rationally* required to be consistent. Insofar as rationality is concerned, we should avoid inconsistent beliefs and believe the necessary implications of what we believe. But rationality is not all that matters. Rational requirements can be 'overridden.' If there is some stronger requirement against being consistent (or against believing the truth and nothing but the truth), then the rational requirement does not generate an all-things-considered requirement. All-things-considered we ought not to be consistent in some situations, although we always ought to be consistent *insofar as* rationality is concerned. So, this counterargument against the abductive argument is not conclusive.

Another possible problem with the abductive argument is that there may be other maxims that explain the consistency norm. The consistency norm can, for example, also be explained by the following principle:

(P). The probability principle. You ought to believe that A iff A is more likely than not-A.

And there are, in principle, countless other norms that can be used to derive the consistency norm. A critic could argue that we must exclude all other possible explanations before we can conclude that we have good reason to believe in the truth norm. Therefore, the abductive argument fails.

However, it is always impossible to investigate all possible explanations of a certain principle. So, if we have to exclude all other possible explanations before

we can conclude that we have good reason to believe in a certain explanation of some principle, then all abductive arguments fail. Accordingly, a defender of the abductive argument could respond that the abductive argument shows that we have good reason to believe in the truth norm, *unless* the critic can find some other, more plausible, norm that can be used to explain the consistency norm. And it is not obvious that anyone has done that. The probability principle, for example, is intuitively plausible. But it is not obvious that it is a better explanation of the consistency norm than the truth norm. And there are several potential problems with this principle. If (P) is true, then the following principle is, for example, not 'valid.' If we ought to believe A and we ought to believe B, then we ought to believe A and B. But intuitively, this principle seems to be reasonable. Obviously, it is not feasible to discuss all more or less plausible explanations of the consistency norm in this paper.

On the one hand, our last counterargument strongly suggests that the abductive argument doesn't prove that the truth norm is 'valid.' On the other, it is not a clear refutation of the abductive argument either.

4. Conclusion

In this paper, I have investigated one version of the so-called truth norm, the principle that we ought to believe that A if and only if A is true. I have interpreted this norm as a wide scope rational requirement. I have discussed two arguments for this version of the norm: 'the extrapolation argument' and 'the abductive argument'. According to the extrapolation argument, we 'derive' the truth norm from a reflection on what it means to be a perfect believer. According to the abductive argument, the truth norm is supported by the fact that it can be used to derive many other plausible doxastic norms. If the abductive argument is sound, the truth norm can be conceived as *the* fundamental norm of (theoretical) rationality (or wisdom). We have seen that the arguments give some support to the truth norm. However, they are not decisive. I have considered some counterarguments. In my opinion, the counterarguments do not refute the arguments for the truth norm, even though they are interesting. So, is it reasonable or not (all-things-considered) to accept the version of the truth norm that we have discussed? Ought we to believe the truth and nothing but the truth? Before we can answer these questions, I believe, we should also discuss various arguments against (WTN). Can it, for example, handle the no-guidance argument that seems to be a problem for the narrow scope version of the truth norm (Glüer and Wikforss 2009, 2010, 2013)? It seems that we sometimes should suspend our judgement (Atkinson 2021). Is this idea consistent with (WTN)? Etc. However, the arguments for (WTN)

in this paper clearly suggest that this principle is worth taking seriously. I conclude that the version of the truth norm that we have explored deserves further investigation.

References

- Atkinson, C. J. 2021. "The aim of belief and suspended belief." *Philosophical Psychology*, 34(4): 581–606.
- Boghossian, P. A. 2003. "The normativity of content." *Philosophical Issues*, 13: 31–45.
- Bykvist, K. and A. Hattiangadi. 2007. "Does thought imply ought?" *Analysis*, 67: 277–285.
- . 2013. "Belief, truth, and blindspots." In *The Aim of Belief*, ed. T. Chan, 100–122. Oxford: Oxford University Press.
- Chan, T. ed. 2013. *The Aim of Belief*. Oxford: Oxford University Press.
- Côté-Bouchard, C. 2016. "Can the aim of belief ground epistemic normativity?" *Philosophical Studies*, 173(12): 3181–3198.
- Engel, P. 2007. "Belief and normativity." *Disputatio*, 2(23): 197–202.
- . 2013. "In defence of Normativism about the aim of belief." In *The Aim of Belief*, ed. T. Chan, 32–63. Oxford: Oxford University Press.
- Fassio, D. 2011. "Belief, correctness and normativity." *Logique Et Analyse*, 54(216): 471–486.
- . 2022. "The Aim of Belief." In *Internet Encyclopedia of Philosophy*. (Autumn 2022 edn.), eds. J. Fieser and B. Dowden. <<https://iep.utm.edu/aim-of-belief/>>.
- Garson, J. 2018. "Modal Logic." In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta. <https://plato.stanford.edu/entries/logic-modal/>.
- Gensler, H. J. 1986. "Ethics is Based on Rationality." *The Journal of Value Inquiry*, 20: 251–264.
- . 1996. *Formal Ethics*. London and New York: Routledge.
- Gibbard, A. 2003. "Thoughts and norms." *Philosophical Issues*, 13: 83–98.
- . 2005. "Truth and correct belief." *Philosophical Issues*, 15: 338–350.
- Glüer, K. and Å. Wikforss. 2009. "Against Content Normativity." *Mind*, 118: 31–70.
- . 2010. "The Truth Norm and Guidance: a Reply to Steglich-Petersen." *Mind*, 119: 757–761.
- . 2013. "Against belief normativity." In *The Aim of Belief*, ed. T. Chan, 80–99. Oxford: Oxford University Press.
- Harman, G. 1986. *Change in View*. Cambridge: MIT Press.

- McHugh, C. 2011. "What do we aim at when we believe?" *Dialectica*, 65: 327–482.
- . 2012. "The truth norm of belief." *Pacific Philosophical Quarterly*, 93(1): 8–30.
- McHugh, C. and D. Whiting. 2014. "The Normativity of Belief." *Analysis*, 74(4): 698–713.
- McNamara, P. and F. Van De Putte. 2021. "Deontic Logic." In *The Stanford Encyclopedia of Philosophy*, ed. E. N. Zalta, <https://plato.stanford.edu/entries/logic-deontic/>.
- Owens, D. 2003. "Does Belief Have an Aim?" *Philosophical Studies*, 115(3): 283–305.
- Papineau, D. 2013. "There are no norms of belief." In *The Aim of Belief*, ed. T. Chan, 64–79. Oxford: Oxford University Press.
- Raleigh, T. 2013. "Belief norms and blindspots." *Southern Journal of Philosophy*, 51(2): 243–269.
- Reisner, A. 2009. "The Possibility of Pragmatic Reasons for Belief and the Wrong Kind of Reasons Problem." *Philosophical Studies*, 145(2): 257–272.
- Shah, N. 2003. "How truth governs belief." *Philosophical Review*, 112: 447–482.
- Shah, N. and J. D. Velleman. 2005. "Doxastic Deliberation." *Philosophical Review*, 114: 497–534.
- Toribio, J. 2013. "Is there an 'ought' in belief?" *Teorema*, 32(3): 75–90.
- Vahid, H. 2006. "Aiming at truth: Doxastic vs. epistemic goals." *Philosophical Studies*, 131(2): 303–335.
- Velleman, D. 2000. *The Possibility of Practical Reason*. Oxford: Oxford University Press.
- Wedgwood, R. 2002. "The aim of belief." *Philosophical Perspectives*, 16: 267–297.
- . 2007. *The Nature of Normativity*. Oxford: Oxford University Press.
- . 2013. "The right thing to believe." In *The Aim of Belief*, ed. T. Chan, 123–39. Oxford: Oxford University Press.
- Steglich-Petersen, A. 2006. "No Norm Needed: On the Aim of Belief." *The Philosophical Quarterly*, 56(225): 499–516.
- . 2009. "Weighing the Aim of Belief." *Philosophical Studies*, 145(3): 395–405.
- . 2013. "Transparency, Doxastic Norms, and the Aim of Belief." *Teorema*, 32(3): 59–74.
- Whiting, D. 2010. "Should I Believe the Truth?" *Dialectica*, 64(2): 213–224.
- . 2012. "Does Belief Aim (Only) at the Truth?" *Pacific Philosophical Quarterly* 93: 279–300.
- . 2013. "Truth: the Aim and Norm of Belief." *Teorema*, 32(3): 121–135.

Daniel Rønnedal

Yamada, M. 2012. "Taking aim at the truth." *Philosophical Studies*, 157(1): 47–59.

Zalabardo, J. 2010. "Why believe the truth?" *Philosophical Explorations*, 13: 1–21.

Zangwill, N. 2005. "The normativity of the mental." *Philosophical Explorations*, 8: 1–19.

DISCUSSION NOTES/ DEBATE

SELF-FAVORING THEORIES AND THE BIAS ARGUMENT

Bálint BÉKEFI

ABSTRACT In a recent article, Bernáth and Tózsér (2021) defend what they call the Bias Argument, a new skeptical argument from expert peer disagreement. They argue that the best contrastive causal explanation for disagreement among leading experts in philosophy is that they adopt their positions in a biased way. But if the leading experts are biased, non-experts either are also biased or only avoid bias through epistemic inferiority. Recognizing this is expected to prompt one to decrease one's confidence in one's philosophical beliefs. This paper argues that some beliefs are immune to a key premise of the Bias Argument. To show this, the paper develops the concepts of self-favoring theories, decisive support, and standing-incommensurable disagreements. A plausible example of a self-favoring theory, dubbed Mere Reformed Protestantism, is sketched. Many disagreements over self-favoring theories and over beliefs decisively supported by self-favoring theories are shown to be standing-incommensurable. It is then argued that when non-experts are in standing-incommensurable disagreements with experts, the standards of assessing expertise are themselves controverted. This result undercuts the move in the Bias Argument from expert bias to non-expert bias. Finally, a couple reservations about the role of self-favoring theories in philosophy are addressed.

KEYWORDS: the Bias Argument, expert peer disagreement, philosophy, self-favoring theories, Bernáth, Tózsér

Introduction

The Bias Argument proposed by Bernáth and Tózsér (2021) seeks to show that philosophers ought significantly to decrease their confidence in their “substantive, positive, and factual philosophical beliefs” (372)¹ upon realizing that these are subject to disagreement among experts, and that the best explanation of this fact is that philosophers are motivated by arational biases in choosing their positions. The authors shirk away from accepting this argument, realizing that it is plausibly self-defeating, but they state that they “know no plausible way to refute its premises” (363). The argument runs as follows (364–365):

(1) In debates about the truth of a given philosophical proposition *p*, those participants who are recognized as the leading experts in the field are *epistemic peers to each other* (with respect to *p*), even though they have conflicting beliefs

¹ Page number-only references refer to Bernáth and Tózsér (2021).

about the truth of p .

(2) If the conflicting philosophical beliefs of epistemic peers have contrastive causal explanations, then these beliefs are biased—their formation and persistence are decisively influenced by factors that do not indicate the truth of these beliefs.

(3) All beliefs (including philosophical ones) have contrastive causal explanations...

(C1) The conflicting philosophical beliefs of the leading experts in philosophy are biased.

(4) If the conflicting philosophical beliefs of the leading experts are biased, then my philosophical beliefs that are heavily debated among experts are biased, too...

(C2) My philosophical beliefs that are heavily debated among experts are biased...

(5) If my philosophical beliefs that are heavily debated among experts are biased, then I have a strong reason to significantly reduce my confidence in them...

(C3) I have a strong reason to significantly reduce my confidence in my philosophical beliefs that are heavily debated among experts.

In response, I argue that one need not accept (4) if one's view under dispute implies that one is epistemically superior to the so-called "leading experts." I call disagreements with this feature standing-incommensurable. In the first section, I show how standing-incommensurable disagreements (SIDs) arise from beliefs decisively supported by self-favoring theories (SFTs). While doing so, I attempt to give a plausible characterization of SFTs and develop an example. In the second section, I consider whether people in SIDs should doubt (1)—and conclude that shouldn't—and elaborate on why they have no reason to accept (4). Finally, I address two reservations one might have about attributing such philosophical relevance to SFTs.

1. How Self-favoring Theories Give Rise to Standing Incommensurability

This section proceeds in four parts. First, Bernáth and Tőzsér's notion of epistemic peerhood is extended to develop an analogous notion of epistemic superiority. Second, this notion is used to characterize SFTs, and an example is introduced. Third, the ways a belief might be "decisively supported" by an SFT are explored. Fourth, it is argued that disagreements over SFTs and many disagreements over beliefs decisively supported by SFTs are standing-incommensurable.

1.1. Epistemic Superiority

Bernáth and Tőzsér define epistemic position or standing in terms of *explanation*: their skeptical conclusion follows when differences in epistemic position cannot

explain a disagreement. (The authors later clarify that by explanation they mean *contrastive causal explanation* (368).) This approach gives rise to the following, original, definition of epistemic peerhood:

Two epistemic agents are epistemic peers to each other in relation to the truth of p if and only if... the difference between their evidence bases, their competences, and their resistance to biases fails to explain their disagreement about the truth of p . (366)

This allows us to propose a parallel definition of epistemic superiority:

Given epistemic agents S and V , S is an epistemic superior of V in relation to the truth of p if and only if (1) S has, on balance, better evidence base, competences, and strength of resistance to biases in relation to the truth of p than V , and (2) their disagreement about the truth of p is explained by (1).

1.2. Self-favoring Theories

This section develops the point that some theories claim or imply that some who accept them are epistemic superiors to all who reject them. Consider *Mere Reformed Protestantism* (MRP), a religious theory. It includes theism, the death and resurrection of Christ, life after death, biblical inspiration, and perhaps some other commitments. It also renders very plausible the following two theses:

The noetic effects of sin. Everyone is sinful; moreover, sin has affected the human epistemic faculties such that all non-regenerate people have a diminished capacity for acquiring and correctly assessing the evidence for MRP.

Regeneration. God has supernaturally restored, at least partially, those epistemic faculties of some people which are relevant for acquiring and correctly assessing the evidence for MRP; moreover, this set of regenerate people is coextensive with those who affirm MRP in some independently identifiable way (the “right way”).²

(The “right way” is left unspecified; it might rule out obviously unreliable methods or postulate a special requirement, such as having undergone a conversion experience.)

It seems plausible that if S accepts MRP, recognizes the conditional plausibility of these two theses, and takes themselves to accept MRP in the right way, then S will rationally come to believe that S is in a better epistemic position concerning the truth of MRP to all people who reject it. S might also rightly think that this difference in epistemic positions contrastively explains the disagreement over MRP. This leads us to the definition of a SFT:

A theory is self-favoring if and only if it implies that an independently

² For a contemporary development of these theses see Plantinga (2000, Ch. 7–8).

identifiable subclass of those who affirm it (“proper believers”) are epistemic superiors in relation to the truth of the theory to all who reject the theory.

One might suspect that SFTs are inherently implausible: how could an uninformed and unintelligent proper believer in a SFT be an epistemic superior to a well-informed and intelligent critic? If SFTs explained disbelief in terms of a systemic lack of communicable information or a systemic prevalence of identifiable errors in reasoning, SFTs would either have to be universally persuasive or become quickly falsified. Instead, there seem to be two viable possibilities. First, a SFT may involve a commitment to a non-standard epistemic principle—a claim that some doxastic method or practice is epistemically reliable (Lynch 2010, 264)—which in turn confirms the SFT. Second, a SFT might be such that disagreements over it must be reducible to disagreements over *basic* doxastic attitudes: basic beliefs, intuitive judgments of plausibility that cannot be further analyzed, or “ur-priors” (Pittard 2019).

Note that this contention is incompatible with some non-explanatory notions of epistemic superiority. Frances, for example, lists a set of individually necessary and jointly sufficient conditions for epistemic superiority, which includes being more informed, intelligent, and intellectually careful, less biased, having thought about the relevant issues longer and more deeply, and having considered the interlocutor’s evidence and reasons (Frances 2010, 420–421). It seems that a proper believer in an SFT might be epistemically superior to someone who rejects the SFT in the explanatory sense—his evidence base and competences are better on balance, and this explain the disagreement—while failing to meet one of Frances’ necessary conditions, such as having thought longer and more deeply about relevant issues than his interlocutor. This is not a problem, however, because the Bias Argument is intimately wedded to explanatory conceptions of epistemic standing.

Notions similar to that of a SFT have appeared in recent literature on the epistemology of religious disagreement. Pittard (2014, 90) talks about *self-favoring theories of epistemic credentials*, which are distinctively *partisan*; Choo (2021, 1143) talks about disagreement over “the relevant credibility-conferring features” which may yield *unconfirmed superiority disagreements*; and Moon (2021) develops the notions of *epistemically self-promoting* and *others-demoting propositions* that allow for implications of unequal epistemic reliability among those who accept and those who reject a given proposition. Compared to these accounts, the two options sketched above shed further light on the possible sources of such epistemic asymmetry.

1.3. Decisive Support

S's belief that *p* is *decisively supported* by *q* if *S* believes that *p* in view of his total evidence *including q* but would *not* believe that *p* if his total evidence *lacked q* but were otherwise identical.³ This notion can be applied to any pair of propositions, but in order to keep its relevance in sight for our purposes, I shall stick to examples involving beliefs decisively supported by MRP. Most trivially, decisively supported beliefs might include propositions that are entailed by *q*—such as theism by MRP. But there may be cases where the addition of *q* tips the scales in favor of *p* even if it does not entail *p*, and even if plenty of other evidence goes into assessing *p*.

Consider substance dualism as an example. There are formidable philosophical and scientific arguments both for and against substance dualism; yet it is probably safe to say that philosophers who lack any religious background are very unlikely to be substance dualists today. Suppose that Doug, an irreligious philosopher, assigns substance dualism an epistemic probability of 0.2 given his knowledge of the relevant evidence. Suppose further that Doug later undergoes a religious conversion and comes to affirm MRP with considerable confidence.⁴ It is perfectly possible that Doug will eventually conclude that the epistemic probability of substance dualism conditional on his original background knowledge in conjunction with MRP is much higher—somewhere around 0.8, say. In such a case Doug would presumably be rational in accepting substance dualism, and this belief of his would be decisively supported by his belief in MRP.⁵

There is at least one further way a belief can be decisively supported by a SFT. Consider Patricia, a moral philosopher who is agnostic about moral realism. Now, Patricia has a specific reason for being agnostic: she has strong moral intuitions which she takes to be *prima facie* evidence for moral realism, but she is also aware of evolutionary debunking arguments against moral intuitions which she takes to undermine any inference from moral intuitions to moral realism. Suppose that upon coming to accept MRP, Patricia realizes that evolutionary debunking arguments now have much less weight, given that she now takes the evolutionary process to have been superintended by a God invested in our moral intuitions. The original defeater for her inference from her moral intuitions to

³ There are a few complications with subtracting a proposition *r* from one's evidence base—issues like conjunctive propositions including *r*, or multiple propositions jointly entailing *r*. Plantinga addresses these complications and eventually suggests: “let's say that $EB_{me}-(B)$ is any subset of EB_{me} [my background evidence] that doesn't entail (B) [the belief to be deleted] and is otherwise maximally similar to EB_{me} ” (Plantinga 2011, 187). This is what is meant here.

⁴ The diachronic nature of these examples only serves illustrative purposes.

⁵ Points along similar lines have been made by Plantinga (1984; 2011, Ch. 6).

moral realism may thus be defeated, and she may rationally come to affirm moral realism—decisively supported by her belief in MRP.

We have thus seen three ways beliefs can be decisively supported by SFTs: entailment, increased conditional probability, and defeater-defeat. Note that these ways can combine and iterate to generate further decisively supported beliefs; however, if a chain of support extends too far, with independent evidence always feeding in, it is likely that the support will eventually cease to be decisive.

1.4. Standing Incommensurability

A disagreement might be said to be a standing-incommensurable disagreement (SID) if the relative epistemic standing of the agents cannot be assessed independently of the truth or falsity of the disputed proposition. I shall argue that disagreements over SFTs and over propositions which for one party are decisively supported by a SFT will often be standing-incommensurable.

Notice that these definitions of epistemic peerhood and superiority in section 1.1 don't say anything about how one might go about assessing epistemic peerhood and superiority. However, the most neutral approach that is also most susceptible to challenges and is hinted at by Bernáth and Tőzsér later (374) is to assess epistemic standing "independently of... views that are closely related to the philosophical issue at hand" (Ibid., emphasis removed).⁶ That is, one ought not reason in the following way: I have very good reason to believe that p ; my interlocutor does not believe that p ; therefore, he just cannot be as bright or well-informed as I am concerning p . Though it is difficult to specify the nuances of this notion of independence (cf. Christensen 2019), the gist of the idea is clear.

A problem emerges, however, when one tries to assess epistemic standing in a disagreement over a SFT. Suppose that "Believer" is a proper believer in such a theory, while "Unbeliever" rejects it. Then, it can be independently recognized that if the theory is true, then Believer is an epistemic superior in relation to its truth to Unbeliever. It should be emphasized that *one need not accept the theory or think that the theory is plausible to accept this conditional*. Suppose further that as far as it can be assessed by ordinary means, Unbeliever is at least as well-informed, intelligent, and (un)biased as Believer. Thus, if the theory is false, then Unbeliever is either Believer's epistemic peer or Believer's epistemic superior.

In such a situation, the relative epistemic standing of "Believer" and "Unbeliever" cannot be decided without taking a position on the truth of the

⁶ In context, the authors are expressing what they take to be the only legitimate way of demoting one's interlocutor.

contested SFT. An independent assessment *yields no result*. It is such cases that are here called SIDs.

Does this just “[push] the problem further down” (373)? Can the standing incommensurability involved in disagreements over SFTs be resolved on the next level? Consider Doug again. Doug and Dan disagree over MRP. Doug believes that this disagreement is explained by Dan’s suffering from the noetic effects of sin, while Dan offers some other explanation. Why do they disagree about the explanation? Both parties may agree that *this* disagreement is explained by Doug’s belief in MRP and Dan lack of such a belief—they have different evidence. Why do they have different evidence? Well, Doug thinks it’s because Dan’s suffering from the noetic effects of sin... and on and on it goes.

More abstractly, the dialectic can be depicted as a circle having two components (Figure 1): (1) disagreement over *the SFT*, and (2) differing explanations of the disagreement. Since Believer’s explanation of their disagreement in terms of Believer’s epistemic superiority is decisively supported by the theory (and Unbeliever presumably rejects this explanation), Believer and Unbeliever’s disagreement over the explanation is explained by their disagreement over the theory (1). No agreement seems to be forthcoming.

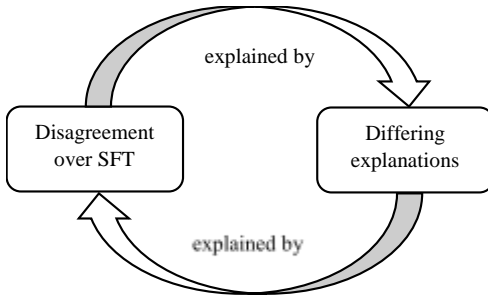


Figure 1: Circularity of explanation in SIDs

In addition to disagreements over SFTs, some disagreements over beliefs decisively supported by SFTs also turn out to be SIDs.⁷ Recall Doug, convert to MRP and substance dualism. Recall further that Doug’s belief in substance dualism is decisively supported by his belief in MRP. Suppose that Dan, an equally sharp and well-informed philosopher, rejects both substance dualism and MRP. Since—by our earlier definition of decisive support—Doug would reject substance dualism if

⁷ That is, if the person appealing to a SFT has reason to consider themselves a proper believer. This qualification will be omitted in subsequent discussions of beliefs decisively supported by SFTs.

he did not accept MRP, their disagreement about substance dualism turns on a disagreement about MRP. Now, it seems that if a disagreement over p turns on a disagreement over q , then the disagreeing parties have the same relative epistemic standing in relation to p as they have in relation to q . But, since MRP is a SFT and Doug is a proper believer in MRP, there is no dispute-independent way to assess their relative epistemic standing—their disagreement is a SID.

Not all disagreements over beliefs decisively supported by SFTs turn out to be SIDs. Consider a modified version of Doug, Dougie: an intelligent layperson who accepts substance dualism and thinks its conditional probability on MRP alone is 0.8. He has a rudimentary understanding of the other evidence, which he judges to support substance dualism to a 0.5 degree. When Dougie learns that Dan, a professional philosopher, rejects substance dualism (and so do most of his colleagues!), he should suspect that there probably is relevant negative evidence that he has not considered and therefore should not take himself to have a better evidence base than the leading experts. To put it another way: because he realizes that he is unfamiliar with Dan's evidence, he is not justified in thinking that their disagreement over MRP contrastively explains their disagreement over substance dualism. Such a situation seems to require two conditions: first, the proper believer needs to think it likely that their interlocutor has weighty arguments and evidence that they are unaware of; and second, the support that the SFT lends to the given philosophical position needs to be moderate, such that other evidence has a realistic chance of offsetting it. Such cases do not generate SIDs.

2. Challenging the Bias Argument

Having introduced SFTs, decisive support, and standing incommensurability, we shall move on to consider the impact of SIDs on the Bias Argument. We will first consider its impact on (1), and conclude that SIDs could in principle, but do not in fact, call it into question. Then we will consider (4) and argue that someone whose belief is decisively supported by a SFT has no reason to accept it, and is thereby immune to the argument.

2.1. Considering (1)

Recall premise (1) of the Bias Argument (365):

- (1) In debates about the truth of a given philosophical proposition p , those participants who are recognized as the leading experts in the field are *epistemic peers to each other* (with respect to p), even though they have conflicting beliefs about the truth of p .

Bernáth and Tózsér defend this premise indirectly (373–374). They argue that the only alternatives to considering the leading experts epistemic peers is to assume one side’s superiority either arbitrarily or at best based on one’s first-order evidence. The authors object that such a move can be mirrored by one’s opponents, and this fact exposes it as unreliable reasoning. The only tenable attitude is to regard experts on both sides of a debate as epistemic peers.

This line of argument seems inapplicable to SIDs, for in such cases, the epistemic superiority of one party is implied by their position. This has two ramifications: first, it is not arbitrary in the sense that Bernáth and Tózsér decry; and second, to regard both parties in SIDs as epistemic peers is not to be neutral or independent, but in fact to claim that the SFT involved is false (for its implication is denied). The only independent assessment would be to suspend judgment about the relative epistemic standing of the disagreeing parties—and then (1) would be unsupported.

But the fact is that most disagreements among leading experts are not SIDs. Most advocates of most or all positions don’t have a SFT decisively supporting their view. Even if 5 out of 10 leading substance dualist experts support their belief decisively by a SFT, that theory is unlikely to explain the disagreement between the *rest* of the leading substance dualists experts and their interlocutors in epistemically relevant terms. Thus, the proper believer may consider their side of the leading experts unbiased only if the disagreement concerns a position which (almost) only proper believers in that theory adopt. Puzzlingly, this becomes more likely the more esoteric a philosophical position is, and the smaller the group of experts defending it is. Even so, it is hard to think of plausible candidates. It seems, then, that since most disagreements among leading experts in philosophy are not SIDs, not even the believer in a SFT has reason to question (1).

2.2. Against (4)

Recall premise (4) of the Bias Argument (365):

- (4) If the conflicting philosophical beliefs of the leading experts are biased, then my philosophical beliefs that are heavily debated among experts are biased, too.

Bernáth and Tózsér motivate this premise by arguing that, assuming that the leading experts’ philosophical beliefs are biased as stated in premise (1), “[t]here are only two things that can save philosophical beliefs from being biased. One is being *uninformed*. ... Another relevant case is *incompetence*” (369, emphasis original). That is to say, if the leading experts are biased but the non-expert’s belief is to be explained by epistemically relevant factors, those factors will involve the non-expert’s epistemic inferiority.

This argument transparently assumes that the leading experts are the epistemic superiors of non-experts. The authors define “the leading experts” as “those philosophers who are more or less consensually regarded by the community of philosophers as the ablest participants in the debate concerning *p*” (365). Thus, the standard of expertise is the consensus of the philosophical community.

At least one party in a SID will deny that this standard of expertise is accurate with respect to the disputed view.⁸ For example, recall Patricia, moral realist and proper believer in MRP. She believes that most philosophers miss a crucial bit of evidence—namely, MRP—and that their missing this bit of evidence contrastively explains their failure to accept moral realism. She also believes that she is in a better position to judge the truth of MRP than those who reject it because of *The noetic effects of sin* and *Regeneration*. Naturally, she also realizes that the consensus of the philosophical community does not take belief in MRP into account when judging who the leading experts are—in fact, she considers it very likely that most philosophers considered to be leading experts do *not* accept MRP. In such a case, she will have no reason to consider the leading experts her epistemic superiors, and thus will not sympathize with Bernáth and Tózsér’s arguments for (4).

To generalize, someone who accepts a “philosophical [belief]... heavily debated among experts” in a way that puts them in a SID with said experts will be unmotivated to accept (4). The Bias Argument does not, therefore, bear on beliefs held in such a way.

3. Two Reservations About Self-favoring Theories

The most likely reaction to the argument up to this point is neither enthusiastic agreement nor outright disagreement. Rather, I expect many readers to be *vaguely unsettled* by this focus on SFTs in philosophy. This section is devoted to addressing a couple possible reservations about belief in SFTs: first, that it is incorrigible; and second, that philosophers should “bracket” it when doing philosophy.

3.1. Is Belief in Self-favoring Theories Incorrigible?

One might worry that SFTs insulate their proper believers into a sort of epistemic circularity by allowing them to reject any objections raised against the theories as resulting from poor evidence, poor competences, or bias. Consequently, proper believers will never have to consider counterevidence to their theories—in a sense,

⁸ For any standard of expertise, at least one party in a SID will deny that it is accurate—otherwise, the standing incommensurability could be resolved.

proper belief in SFTs turns out to be incorrigible (and resulting SIDs irresolvable). If this were so, that would certainly be concerning. As we shall see, however, the objection is hasty. I will show this through considering two distinctions in recent epistemology relevant to theory assessment.

When considering arguments against Christian belief, Alvin Plantinga distinguished *de facto* and *de jure* objections. *De facto* objections, such as many versions of the problem of evil, offer reasons to think that Christian belief is *false*. *De jure* objections, on the other hand, like the Freudian projection theory of theistic belief, argue that Christian belief is *irrational*, *unjustified*, or *unwarranted*—no matter whether true or false (Plantinga 2000, viii–x). Clearly, the objection that one is biased in holding a belief falls in the latter category: the belief's truth value plays no role in the argument. The thesis of this paper, then, is that SFTs are immune to one kind of *de jure* objection—a bias objection from peer disagreement—but they are not, or at least we have seen no reason to think they are, immune to *de facto* objections one might bring against them.

This point is further borne out by considering a similar distinction in recent literature on the epistemology of disagreement: the distinction between first-order evidence and higher-order evidence. First-order evidence E_1 for p is evidence relevant to assessing the truth of p . Higher-order evidence E_H about E_1 for p is evidence relevant to judging how well E_1 's relation to the truth of p has been assessed (cf. Kelly 2016, Section 2). The concept of higher-order evidence is useful for connecting distinct kinds of epistemologically relevant information that seem to threaten with defeat, such as peer disagreement, poor track record, unreliable cognitive faculties, and having taken a rationality-distorting drug (cf. Christensen 2010). On such a schema, this paper argues that SFTs are immune to a certain kind of higher-order evidence (namely, peer disagreement)—but debate concerning the first-order evidence is left completely open to discussion.

We have thus seen that SFTs in principle allow for considering objections and challenging the evidence that purportedly supports them. Two points should nevertheless be acknowledged in favor of the objection. First, it may be the case that one has less rational obligation to consider the objections of one's epistemic inferiors than one's epistemic peers—though, perhaps, the obligation is less diminished in SIDs than in ordinary cases. This is far from incorrigibility, but it does mean that SFTs are more resistant to revision than more trivial beliefs. Second, *some* disagreements over SFTs may turn out to be rationally irresolvable. But this will not be merely in virtue of the theories being self-favoring, but rather some more specific feature, such as their involving contested fundamental epistemic principles (Lynch 2010).

3.2. Should Self-favoring Theories Be Bracketed?

Decisively supporting one's philosophical beliefs by a SFT might seem *unphilosophical*, perhaps illegitimately mixing religion with philosophy (cf. Schellenberg 2018). Maybe we should ask philosophers whose beliefs are decisively supported by SFTs to join those whose are not in a conditional project: they should strive to assess the probability of a given philosophical thesis *conditional on their shared evidence*, ignoring the support some of them might derive from theories not all of them share.⁹ In a sense, they are asked to "bracket" their theories while doing philosophy. Then, if they fail to come to a consensus on this conditional probability, then they should significantly decrease their confidence that their shared evidence supports the given philosophical thesis to that certain degree.

Should the proper believer agree to this? Not always, for her theory might claim that she is in a better epistemic position to assess not just the theory itself, but also other issues close to its core claims. Thus, even if she were to bracket her theory *as evidence*, her theory might still provide a contrastive explanation of the disagreement over the assessment of the agreed-upon evidence in terms of her epistemic superiority. For example, a believer in MRP who accepts *The noetic effects of sin* may well be justified in thinking that those who reject MRP tend to be more biased in their assessment of the evidence for theism (entailed by MRP) than she is. So, if a SFT implies that its proper believers are epistemic superiors to those who reject the theory with respect to assessing the support their shared evidence lends to some philosophical view in domain *D*, then the Bias Argument does not demand that proper believers decrease their confidence in their initial assessment.

But suppose the given SFT does *not* imply its believer's epistemic superiority with respect to some *p* (other than the epistemic superiority gained by having the theory in one's evidence base), belief in which may nevertheless be decisively supported by the theory. Then, the philosopher accepting it may have to accept the Bias Argument and decrease her confidence in the probability of *p* conditional on the evidence shared by the philosophical community. This *would* impact her confidence and, in some cases, even her belief in *p*.

It is worth laying out three further considerations that limit the SFT believer's steadfastness. First, SFTs tend to restrict the claim to epistemic superiority to some limited domain. Second, epistemically relevant differences may explain some amount of disagreement while not explaining all amount. For example, MRP might give an epistemically relevant explanation for why Doug

⁹ Here we diverge from Bernáth and Tőzsér's binary (belief–disbelief) model of disagreement.

would assign a 0.6 probability to the thesis that an infinite past is metaphysically impossible, while Dan would judge it to be 0.4 probable.¹⁰ But MRP might *not* have an explanation in terms of epistemically relevant differences for a disagreement between probability judgments of 0.9 and 0.1. Third, the argument of this paper does not exempt believers in the *same* SFT of the skeptical consequences of disagreement.

4. Conclusion

This paper has argued that there is a special category of beliefs that is immune to Bernáth and Tőzsér's Bias Argument. To describe this category of beliefs, new concepts had to be proposed: those of a SFT, decisive support, and SIDs. It was argued that all disagreements over SFTs between proper believers and disbelievers and many but not all disagreements over beliefs decisively supported by SFTs are SIDs: there is no dispute-independent way to assess the relative epistemic standing of the two parties.

This conclusion was then applied to the Bias Argument. It yielded the conclusion that if the disagreement over some philosophical view between a person and "the leading experts" in philosophy is a SID, then that person has no reason to think that "the leading experts" are their epistemic superiors, and thus no reason to accept premise (4). Moreover, it was shown both through the concept of SFTs and through illustrations that such SIDs are perfectly possible.

The concept of SFTs may provoke objections; therefore, the final section of the paper was devoted to addressing two of these. It was argued that SFTs need to be incorrigible, as standing incommensurability does not prevent the believer in a SFT from examining *de facto* objections (objections concerning first-order evidence) to their view. The suggestion that belief in SFTs should be bracketed while doing philosophical work was shown to be ill-conceived, as it would not secure the success of the Bias Argument. This is because SFTs may claim that their proper believers are epistemic superiors to others not just with respect to the theory itself, but also with respect to other closely related domains of inquiry.

The paper raises questions for future inquiry about how the concepts proposed here relate to other discussions. What is the relationship between SIDs and deep disagreements? What views qualify as SFTs in philosophical, religious,

¹⁰ How might MRP explain this? Most philosophers agree that the finitude of the past provides at least *some* evidence for theism; but, as we have said, *The noetic effects of sin* suggests that those who deny MRP are more biased in their assessment of the evidence for theism than those who accept it in the right way.

perhaps even psychological or economic thought?¹¹ These questions lie beyond the scope of the present work.

References

- Bernáth, László and János Tózsér. 2021. "The Biased Nature of Philosophical Beliefs in the Light of Peer Disagreement." *Metaphilosophy* 52:363–378. <https://doi.org/10.1111/meta.12501>.
- Choo, Frederick. 2021. "The Epistemic Significance of Religious Disagreements: Cases of Unconfirmed Superiority Disagreements." *Topoi* 40:1139–1147. <https://doi.org/10.1007/s11245-018-9599-4>.
- Christensen, David. 2010. "Higher-Order Evidence." *Philosophy and Phenomenological Research* 81 (1):185–215. <https://doi.org/10.1111/j.1933-1592.2010.00366.x>.
- . 2019. "Formulating Independence." In *Higher-Order Evidence: New Essays*, edited by Mattias Skipper and Asbjørn Steglich-Petersen, 13–34. Oxford: Oxford University Press.
- Frances, Bryan. 2010. "The Reflective Epistemic Renegade." *Philosophy and Phenomenological Research* 81 (2):419–463. <https://doi.org/10.1111/j.1933-1592.2010.00372.x>.
- Kelly, Thomas. 2016. "Evidence." In *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), edited by Edward N. Zalta. <https://plato.stanford.edu/archives/win2016/entries/evidence/>.
- Lynch, Michael P. 2010. "Epistemic Circularity and Epistemic Incommensurability." In *Social Epistemology*, edited by Adrian Haddock, Alan Millar, and Duncan Pritchard, 262–277. Oxford: Oxford University Press.
- Moon, Andrew. 2021. "Circular and Question-Begging Responses to Religious Disagreement and Debunking Arguments." *Philosophical Studies* 178:785–809. <https://doi.org/10.1007/s11098-020-01458-3>.
- Pittard, John. 2014. "Conciliationism and Religious Disagreement." In *Challenges to Moral and Religious Belief: Disagreement and Evolution*, edited by Michael Bergmann and Patrick Kain, 80–97. Oxford: Oxford University Press.
- . 2019. "Fundamental Disagreements and the Limits of Instrumentalism." *Synthese* 196:5009–5038. <https://doi.org/10.1007/s11229-018-1691-1>.

¹¹ For a few imaginative non-religious suggestions see Turnbull (2021).

- Plantinga, Alvin. 1984. "Advice to Christian Philosophers." *Faith and Philosophy* 1 (3):253–271. <https://doi.org/10.5840/faithphil19841317>.
- . 2000. *Warranted Christian Belief*. Oxford: Oxford University Press.
- . 2011. *Where the Conflict Really Lies: Science, Religion, and Naturalism*. Oxford: Oxford University Press.
- Schellenberg, J. L. 2018. "Is Plantinga-Style Christian Philosophy Really Philosophy?" In *Christian Philosophy: Conceptions, Continuations, and Challenges*, edited by J. Aaron Simmons, 229–243. Oxford: Oxford University Press.
- Turnbull, M. G. 2021. "Religious Disagreement Is Not Unique." In *Religious Disagreement and Pluralism*, edited by Matthew A. Benton and Jonathan L. Kvanvig, 91–107. Oxford: Oxford University Press.

HAVING A HUNCH

Howard SANKEY

ABSTRACT: It has recently been argued that when one conducts an inquiry into some question one ought to suspend belief with respect to that question. But what about hunches? In this short note, a hunch about the cause of a phenomenon is described. The hunch plays a role in the inquiry into the cause of the phenomenon. It appears that the hunch constitutes a belief that need not be suspended during the inquiry even though belief about the precise cause of the phenomenon is suspended.

KEYWORDS: inquiry, suspension of belief, Jane Friedman, hunches

I.

The other evening while watching the news, I heard a loud noise outside. At first, there was a rustling sound. Then there was a sliding or scraping sound. Finally, there was a loud thud, a crash.

I was curious about the cause of this commotion. So, I went outdoors to investigate. What I found was that a large branch about 15 meters in length had broken off at its base from a neighboring eucalyptus. It had fallen in such a way that the base of the branch was embedded in the ground. The base of the branch was about a third of a meter across. The branch itself was tilted at a 45-degree angle. The top of the branch was firmly wedged in forks of branches of two smaller trees.

In going outside to find out what made the noise, I conducted an inquiry. I did not know what made the noise. I went to have a look. I observed the branch in its new position. As a result, I came to know that it was the falling limb that had made the noise.

I started in a position of ignorance. I conducted an inquiry. I observed the relevant evidence. I found out what caused the noise. The inquiry was complete. At that time, I knew what caused the noise.

Here the process is a process of inquiry which moves from ignorance to knowledge. I start with ignorance. I conduct an inquiry. I wind up with knowledge.

But there is a hitch. I had a hunch it might have been a falling branch.

II.

It was quite a strong hunch. I live in a forest. Eucalypts sometimes drop major branches. It had been hot. Eucalypts are known to drop branches as a result of the heat. So, I thought that it might be a branch.

Still, I wasn't certain. There were other possible explanations. It could have been a drop bear, but they are rare as hen's teeth.¹ An asteroid strike was unlikely. An earthquake? But the house didn't shake. Perhaps a car had gone down the ravine across the road (it has happened before). But there was no sound of skidding tires, smashed fence posts, crumpling metal, breaking glass or human voices.

So, I had a hunch, quite a strong one, that a branch had fallen. But I was not certain. I needed to check. I needed to have a look to see if I could determine the cause. I walked in the direction of the noise and saw the fallen branch. It was just on dark. So, I waited until dawn to check it out in detail.

III.

In an inquiry such as this one, there is a practical component. To see what made the noise, I went outside. Rather than head down the drive, I went up the drive behind the house to look in the paddock next door. That was because I thought the sound came from that direction. I walked in the direction that I thought the noise had come from and I looked toward the paddock once I got to where I could see the paddock.

This shows that inquiry involves, or may involve, a practical component. One engages in a series of actions in order to conduct the inquiry. I went outside and looked toward the paddock to determine the source of the noise. But the bodily movements involved did not in themselves constitute the inquiry (cf. Friedman 2019, 297). I had made similar movements countless times before over the years without necessarily inquiring into anything. It was the fact that I was trying to determine what had made the sound that made these movements into components of an act of inquiry.

Having seen the fallen branch, I knew what caused the noise. Given the evidence, I knew what the cause was. But still I returned the following morning to check it out further. Why? Was I double-checking or re-checking? Well, in a

¹ The drop bear (*Thylarctos plummetus*) is a rare and possibly mythical relative of the koala that inhabits dense eucalypt forests, where it drops from great height onto unsuspecting tourists. See Middleton (2021).

sense, yes, but not to determine the source of the noise. When I went back at first light to check the branch, I was more interested in what to do next. How big was it? How firmly wedged was it? Was it safe? Would I be able to cut it down with my chainsaw? Or would I need to call the arborist?

These were the questions before my mind when I went back to check the tree in the morning. I was not checking again to make sure that it was the branch that had made the crashing sound. I knew what made the crashing sound. But now I had other questions to resolve, ones that related to what I should do next.

IV.

Much of what I have just said is familiar ground from the current literature on the nature of inquiry. Inquiry is a process that moves from ignorance to knowledge. It seeks to answer a question (e.g., what made that noise?). An inquiry has an aim (e.g., to find out the source of the noise). It may involve practical activity (e.g., going to look). Sometimes we double-check, though in this case my double-checking was not strictly double-checking, since it was in service of different questions.

What is not so familiar is the status of the hunch.

According to Jane Friedman (2019, 303), when we conduct an inquiry, we should suspend belief. We do not genuinely conduct an inquiry if we already have a belief about the matter at hand. There is something wrong with conducting an inquiry into some matter if I have already formed a belief with respect to that matter. Thus, I should not have gone out to determine the cause of the noise if I believed or knew that a falling branch had made the noise. If I knew or believed this, I would have had no need to make the inquiry. It would have been inappropriate for me to do so. It is irrational to conduct an inquiry into a matter about which one already believes or knows the answer. It is not irrational to conduct the inquiry if the belief is suspended.

This doesn't sit quite right with me. I did not outright believe that it was a falling branch that had caused the noise. I was not completely confident that the cause was a falling branch. I had made no commitment to that precise belief. I had not made up my mind. That is why I went outside to investigate. But I did have a hunch that it was a falling branch that made the noise.

Is the hunch a belief?

V.

Here is my best stab at what the hunch amounted to. My hunch that a falling branch made the noise was a belief that it was probably a branch that made the

noise. This is not the belief that a falling branch made the noise. That belief was suspended until I saw the fallen branch. Instead, the belief was about the probability or likelihood (we may be loose about terminology here) that it was a falling branch that made the sound. This was not partial belief (less than full confidence) that a falling branch made the sound (that belief was suspended). It was outright belief that it was probable that a falling branch made the sound (that belief was not suspended).

There is a (physical) context. I live in a forest. The trees among which I live often drop branches. On hot days, eucalypts sometimes drop branches. I am familiar with many of the noises that occur around the house. I know many of the bird calls. I know what wind in the trees sounds like. I know the sounds that come from the neighboring properties and the road. Most importantly, I have heard the sound that branches make when they fall on previous occasions.

That context informs my belief-formation. The beliefs that I form in the circumstances in which I find myself are guided by beliefs and knowledge that I have about my surroundings. It's because I live in a forest and am aware of what goes on around me that I formed the belief that it was likely that a falling branch had made the noise.

That belief, the belief that it was likely that a falling branch made the noise, was the hunch that I had about the cause of the noise.

What role did the hunch play? I had not yet formed the belief that the falling branch made the noise. But the hunch that a branch made the noise served as a guide to my inquiry. The inquiry led to the belief that the falling branch made the noise. But the inquiry itself was guided by the hunch. It was because of the hunch that I went outside to investigate. It was the hunch that prompted and guided the actions that I undertook to make the inquiry.

In what I have just said, I have written in descriptive mode. I have described the way in which the surroundings in which I live shaped my hunch about the falling branch. But Friedman's point about suspension of belief in inquiry is a normative one. So, the description of the circumstances that inform my hunch may seem to miss the point. I do not think that it does. It seems entirely appropriate, indeed rational, for those circumstances to inform my hunch and for the hunch to guide my inquiry.

VI.

The suggestion I venture to offer is that in at least some cases hunches play and rightly play a role in our inquiries. We may suspend belief about the precise matter into which we inquire, as I did about whether a falling branch made the noise. But

if we inquire based on a hunch, as I did into the noise the other evening, our inquiry is and should be guided by belief. In this case, my inquiry was guided by the belief that it was probably a falling branch that made the noise. And that was entirely appropriate from a rational point of view.

Inquiries made based on hunches are not inquiries in which all relevant beliefs are suspended. Nor should they be.

References

- Friedman, Jane. 2019. "Inquiry and Belief." *Nous* 53: 2, 296-315
- Middleton, Amy. 2021. "Drop bears target tourists, study says." *Australian Geographic* 1/04/2021, <https://www.australiangeographic.com.au/news/2021/04/drop-bears-target-tourists-study-says/>

DISTRACTIONS

RADICAL KNOWLEDGE MINIMALISM

Jeremy FANTL, Matthew McGRATH

ABSTRACT: We argue that knowledge doesn't require any of truth, justification, or belief. This is so for four primary reasons. First, each of the three conditions has been subject to convincing counterexamples. In addition, the resultant account explains the value of knowledge, manifests important theoretical virtues (in particular, simplicity), and avoids commitment to skepticism.

KEYWORDS: knowledge minimalism, truth, justification, belief, skepticism

In this short paper we argue that the JTB account of knowledge is false in all its components: knowledge doesn't require any of truth, justification, or belief. You can know that p whether p is true or false, whether p is justified or not, and whether you believe that p or don't. Indeed, knowledge does not require the satisfaction of any substantive conditions whatsoever. In short, you know **Whatever** is **True** or **False**. Call this the WTF account of knowledge:

(WTF) S knows that p iff p or not- p .

There are four primary arguments for WTF. First, convincing arguments have been mustered against each of the three conditions in the traditional JTB account, and it is time we took those arguments seriously and jointly. Second, WTF explains the value of knowledge. Third, WTF manifests important theoretical virtues. Fourth, WTF avoids what is widely agreed to be an implausible commitment to skepticism. In this paper we present these four arguments and then consider three objections to our project.¹

Argument 1: The Existing Literature

Allen Hazlett (2010) has argued that knowledge doesn't require truth. Crispin Sartwell ((1991) and (1992)) has argued that knowledge doesn't require justification. Colin Radford (1966) has argued that knowledge doesn't require

¹ If you're worried that the WTF account is too disjunctive, note that it is roughly equivalent to a non-disjunctive account according to which you know that p iff p has a truth value. This non-disjunctive account allows for truth values other than truth or falsity, but this may even seem to be a virtue of the non-disjunctive account. We leave this issue for more extensive discussions of the WTF account.

Jeremy Fantl, Matthew McGrath

belief. They are all correct. Therefore, knowledge doesn't require truth, justification, or belief.

This argument is further strengthened by noting that each view was plausible enough to be published. But consider the following Agglomeration Thesis:

(Agglomeration) If A was plausible enough to be published and B was plausible enough to be published and C was plausible enough to be published, then A&B&C is plausible enough to be published.

It follows from Agglomeration and the fact that the denial of each of the three JTB conditions was plausible enough to be published, that the conjunction of the denials of the three JTB conditions is plausible enough to be published. We leave the derivation of Agglomeration to the reader.

Argument 2: The Value of Knowledge

In the *Meno*, Plato tentatively suggests that knowledge is especially valuable because it is a particularly stable cognitive attitude. Unlike mere true belief, which is not tied down, knowledge lasts. Timothy Williamson, more recently, has agreed that knowledge is "robust" with respect to "destruction by later evidence" (Williamson 2000 63). When you know something, it is resistant to misleading counterevidence.

The WTF account allows knowledge to be *radically* stable. Not only is knowledge resistant to misleading counterevidence, it survives any evidence entirely – misleading or not.

Argument 3: Theoretical Virtues of WTF

It is a virtue of a theory that it be simple. There is no other theory of knowledge as simple as WTF.

Argument 4: Anti-Skeptical Consequences of WTF

Skepticism is often assumed to be a deal-breaker in contemporary epistemology. Laurence Bonjour refers to very strong forms of skepticism as "intellectual suicide" (1998, 5) while David Lewis is even willing to choose fallibilism over the "whirlpool of skepticism" (1996, 550). The primary worry about skepticism is that, if it's true, we don't know anything we take ourselves to know.

WTF has no skeptical consequences. On WTF, you know everything you take yourself to know. You know everything else, as well. WTF is maximally anti-skeptical.

Objections

Objection 1: Self-defeat

The first objection is an *ad hominem*: it's that we – the authors of this paper – do not even believe our conclusion and so, by our own lights, shouldn't be asserting it.

Reply: the norm of assertion is knowledge, not belief. By our own lights (that is, according to WTF), belief is not necessary for knowledge (and nor is truth). WTF itself is either true or false. Therefore, by our own lights, we know that WTF is true. Therefore, by our own lights, we may assert that WTF is true.

Objection 2: Gettier Counterexamples

The second objection directs Gettier counterexamples against WTF. In Gettier counterexamples, your failure to know is not attributable to a failure of belief, truth, or justification. Mustn't, then, there be an additional condition on knowledge – one that we, in arguing for WTF, have not yet ruled out?

Reply: Gettier cases are cases in which you satisfy some conditions on knowledge only as a result of luck. But on WTF, you never only luckily satisfy any conditions on knowledge; you never could have easily failed to satisfy any of the conditions on knowledge. It can't easily be the case that neither p or not- p in possible worlds in which p or not- p . Of course, there are some situations in which you could easily have believed falsely. But because knowledge doesn't require either belief or truth, it is irrelevant that you could have easily believed falsely. Knowledge doesn't require non-lucky satisfaction of conditions that are irrelevant to knowledge.

Objection 3: Traditional Counterexamples

The third objection points to various cases in which it seems that someone fails to know something, whether because what they know is false, they fail to believe it, or they fail to have justification. The WTF account predicts that no one ever fails to know anything. Therefore, any case in which someone fails to know something is a counterexample to the WTF account.

Reply: Sartwell points out that the case for eliminating the justification condition “cannot be refuted by the flick of a counterexample” (1992, 167). The point generalizes.


Of course, we have relied, in small part, on at least one author who resists traditional conditions on knowledge by reference to counterexamples. But all this shows is that all competitor accounts are subject to intuitive counterexamples. We're left to decide among the competitor accounts according to how well they do

on a checklist of desiderata. As argued above, the WTF account does exceedingly well on anti-skeptical criteria, simplicity (and, so, theoretical virtues), and explaining the value of knowledge. It also preserves closure, explains why knowledge is a necessary condition for proper assertion, allows infants and animals to have knowledge, does not require an ideal theory in epistemology, and does not overly intellectualize knowledge. No competitor account does as well on all of these criteria.

Brian Weatherson says that a theory can be correct if it does well enough on a variety of theoretical tests to overcome intuitive difficulties with counterexamples. (2003, 10) We concur. This is especially so if all the rival accounts face their own intuitive difficulties. That there may be occasional criteria on which the WTF account fares worse does not alter the fact that the WTF account should be considered an important contender on the epistemological stage. The relative intuitive plausibility of rival accounts is just one small factor among many.

Conclusion

It might be thought that WTF isn't minimal enough. Why restrict knowledge to claims that are either true or false? What of the neither true nor false? Why allow this chauvinism? Should WTF be replaced with the view that S knows whatever is either true or false or neither true nor false?

While we are sympathetic to the possibility of knowledge of the neither true nor false, we think the consequences are implausible. Rocks, words, and incoherent phrases are all neither true nor false, so expanding the domain of knowledge in this way allows knowledge that desktop, knowledge that blue prime number, and knowledge that . This is absurd. As Weatherson notes, "While a theory can be reformist, it cannot be revolutionary" (2003, 8). Therefore, you know all and only what is true or false.²

References

BonJour, Laurence. 1998. *In Defense of Pure Reason*. Cambridge, Cambridge University Press.

² We don't have the resources to exclude an account of knowledge according to which you know that p iff p is both true and false. Though the simplicity of this account, not to mention the fact that this account explains why knowledge is a rare jewel indeed, tempts us, the complete argument against it would take us beyond the scope of this paper. We leave the maximalist account for future researchers.

- Hazlett, Allan. 2010. "The Myth of Factive Verbs." *Philosophy and Phenomenological Research* 80(3): 497-522.
- Lewis, David. 1996. "Elusive Knowledge." *Australasian Journal of Philosophy* 74: 549-67.
- Radford, Colin. 1966. "Knowledge—By Examples." *Analysis* 27(1-11).
- Sartwell, Crispin. 1991. "Knowledge Is Merely True Belief." *American Philosophical Quarterly* 28(2): 157-65.
- Sartwell, Crispin. 1992. "Why Knowledge Is Merely True Belief." *The Journal of Philosophy* 89(4): 167-80.
- Weatherson, Brian. 2003. "What Good Are Counterexamples?" *Philosophical Studies* 115: 1-31.
- Williamson, Timothy. 2000. *Knowledge and Its Limits*. Oxford, Oxford University Press.

NOTES ON THE CONTRIBUTORS

Miguel López Astorga has a Ph.D. in Logic and Philosophy of Science from the University of Cádiz (Spain). He carried out post-doctoral research on human reasoning, on the mental model theory, and on the logical forms of propositions, under the supervision of Professor Philip Johnson-Laird. Although he has worked in several institutions, at present, he is Full Professor and Researcher at the Institute of Humanistic Studies, Research Center on Cognitive Sciences, University of Talca (Chile). His last researches focus on the human inferential activity. He is especially interested in the relations that can be found between syntax, semantics, and pragmatics in that activity. He also addresses the possible cognitive potential of some ancient logics, particularly, Stoic logic. ORCID ID: <https://orcid.org/0000-0002-6004-0587>. Contact: milopez@utalca.cl.

Bálint Békefi is a graduate student in Philosophy at Central European University and an Associate Lecturer at Pentecostal Theological College, Budapest. His main areas of research are epistemology and the philosophy of religion. Contact: balint.bekefi@gmail.com.

Jeremy Fantl is Professor of Philosophy at the University of Calgary. He is the author of *The Limitations of the Open Mind* and, with Matthew McGrath, *Knowledge in an Uncertain World*. He works primarily in epistemology. Contact: jfantl@ucalgary.ca.

Ali Hossein Khani is Researcher at University of Religions and Denominations (Tehran) and Assistant Professor at Iranian Institute of Philosophy (Tehran). His publications have been mainly on Quine, Davidson, Wittgenstein and Wright. He is the author of the monograph on Kripke, *Kripke's Wittgenstein: Meaning, Rules, and Scepticism* (Anthem Press, forthcoming), and the co-editor, with Gary N. Kemp, of two volumes on Wittgenstein, *Wittgenstein and Other Philosophers, Volume I and II* (Routledge, forthcoming) as well as a collected volume on naturalism, *Naturalism and Its Challenges* (Routledge, forthcoming). ORCID: <https://orcid.org/0000-0002-0537-3805>. Contact: hosseinkhani@irip.ac.ir.

Jimmy Alfonso Licon is an Assistant Teaching Professor of Philosophy at Arizona State University. He received his Ph.D. in philosophy from the University of Maryland, College Park. His research specializes in ethical and epistemic aspects of

cooperation, signaling, and ignorance, and the role of aspirations and evidence in philosophy of religion and philosophy of law. Contact: jimmy.licon@asu.edu.

Matthew McGrath, Professor of Philosophy at Washington University in St. Louis, is the author, with Jeremy Fantl, of *Knowledge in an Uncertain World* as well as a number of papers in epistemology and metaphysics. Contact: matthew.m@wustl.edu.

Daniel Rönndal is a Swedish philosopher, logician and researcher affiliated with Stockholm University. He has published papers on modal logic, doxastic logic, deontic logic, temporal logic, counterfactual logic, various paradoxes and aporias (such as the contrary-to-duty paradoxes), the golden rule and several other topics. Contact: daniel.ronnedal@philosophy.su.se.

Howard Sankey is Professor of Philosophy in the School of Historical and Philosophical Studies at the University of Melbourne (Australia). He teaches in epistemology and philosophy of science. He has published on the problem of incommensurability, epistemic relativism and scientific realism. He is the author of *Scientific Realism and the Rationality of Science* (Ashgate, 2008), *Theories of Scientific Method: An Introduction* (Acumen, 2007, with Robert Nola), *Rationality, Relativism and Incommensurability* (Ashgate, 1997) and *The Incommensurability Thesis* (Avebury, 1994). More information may be found at <https://philpeople.org/profiles/howard-sankey>. Contact: chs@unimelb.edu.au.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English which satisfy the norms of clarity and rigour in exposition and argumentation.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the **Notes to Contributors**.

Contact: logosandepisteme@yahoo.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com.

Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*.rtf) or Microsoft Word document (*.docx) and must be double-spaced, including quotes and footnotes, in

12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author's affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in

their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.

11. Checking Proofs


Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice și Sociale „Gh. Zane” Academia Română, Filiala Iași, Str. Teodor Codrescu, Nr. 2, 700481, Iași, România.

The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Copyright & Publishing Rights

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License (). Authors have the right to use, reuse and build upon their papers for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme* are also freely available on Philosophy Documentation Center web page.