

Volume XI ♦ Issue 3

2020

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

Alexandru Surdu

Academia Română, România

John F. Symons

University of Texas at El Paso, USA

TABLE OF CONTENTS

RESEARCH ARTICLES

Robb DUNPHY, Agrippan Problems.....	259
Domingos FARIA, Group Testimony: Defending a Reductionist View.....	283
Thomas GRUNDMANN, Moral Realism and the Problem of Moral Aliens.....	305
Perry HENDRICKS, The Subject's Perspective Objection to Externalism and Why it Fails.....	323
Robert MICHELS, Husserlian Eidetic Variation and Objectual Understanding as a Basis for an Epistemology of Essence.....	333
Diana SOFRONIEVA, Empathy as a Tool for Learning about Evaluative Features of Objects.....	355
Billy WHEELER, Truth Tracking and Knowledge from Virtual Reality.....	369

DISCUSSION NOTES/ DEBATE

Scott AIKIN, Does Metaphilosophically Pragmatist Anti-Skepticism Work?....	391
Notes on the Contributors.....	399
<i>Logos and Episteme</i> . Aims and Scope.....	401
Notes to Contributors.....	403

RESEARCH ARTICLES

AGRIPPAN PROBLEMS

Robb DUNPHY

ABSTRACT: In this article I consider Sextus' account of the Five Modes and of the Two Modes in his *Outlines of Pyrrhonism*. I suggest that from these we can derive the basic form of a number of different problems which I refer to as "Agrippan problems," where this category includes both the epistemic regress problem and the problem of the criterion. Finally, I suggest that there is a distinctive Agrippan problem present at the beginning of Hegel's *Science of Logic*.

KEYWORDS: Pyrrhonian Scepticism, Modes, epistemic regress problem, Hegel

1. Introduction

In his *Outlines of Pyrrhonism*, Sextus attributes two sets of modes to "the more recent Sceptics" (*PH* I: 164).¹ These are the Five Modes (*PH* I: 164-77) and the Two Modes (*PH* I: 178-79). In his account of Pyrrhonian Scepticism, Diogenes Laërtius attributes the Five Modes to a Sceptic² called Agrippa.³ Accordingly, it is common to refer to these as the 'Agrippan modes.' Since it is not unusual firstly to hold that the Two Modes amount to a compact restatement of the problem expressed by way of the Five Modes,⁴ and secondly to attribute the Two to the same author as that of the Five,⁵ we can also refer to the Two Modes as 'Agrippan modes.' Ultimately, the

¹ *PH* = Sextus Empiricus, *Outlines of Scepticism*, trans. J. Annas and J. Barnes (Cambridge: Cambridge University Press, 2000). Reference is provided to book and Fabrician section number. Translations have occasionally been altered.

² I will follow common practice and capitalise 'Sceptic' when referring to the position of the Pyrrhonian sceptic as portrayed in the works of Sextus, and capitalise 'Dogmatist' when referring to their opponents, in response to whose philosophical claims the Sceptics attempted to prompt a suspension of judgement.

³ Diogenes Laërtius, *Lives of the Eminent Philosophers*, trans. P. Mensch (Oxford: Oxford University Press, 2018), IX: 89

⁴ See, for example, J. Barnes, *The Toils of Scepticism* (Cambridge: Cambridge University Press, 1990): 117; R.J. Hankinson, *The Sceptics* (New York, NY: Routledge, 1995): 189; G.W.F. Hegel, "On the Relationship of Skepticism to Philosophy, Exposition of its Different Modifications and Comparison of the Latest Form with the Ancient One," in *Between Kant and Hegel: Tests in the Development of Post-Kantian Idealism*, eds. G. di Giovanni and H.S. Harris (Indianapolis, IN: Hackett, 2000): 335; R. Bett, *How to be a Pyrrhonist: The Practice and Significance of Pyrrhonian Scepticism* (Cambridge: Cambridge University Press, 2019): 108-09.

⁵ See, for example, Barnes, *The Toils of Scepticism*: 117, n.3; T. Brennan and J. Lee, "A Relative

attribution of these modes to Agrippa is not terribly important, since none of his works survive and we know nothing about him beyond his brief mention in Diogenes' account of the Pyrrhonian tradition, but it provides a useful label with which to group together the sceptical material under examination here.

I shall argue that the famous trilemma embedded in the Five Modes and its more compact expression as a dilemma in the Two Modes provide us with the basic form of philosophical problems which can be called 'Agrippan problems.' These include the 'epistemic regress problem,' the 'problem of the criterion,' Hegel's peculiar 'problem of beginning,' and others besides, although it is not my intention to provide an exhaustive list. I aim to identify the defining characteristics, then, of a fairly extensive family of problems. I suggest that, at their simplest, these problems confront us with a dilemma between the arbitrariness of a claim put forward without any support and the arbitrariness of a claim put forward on the basis of fundamentally inadequate support. This essay, which will involve a certain amount of jumping around in the history of epistemology, is primarily exploratory, but I hope that it can contribute to laying the foundations for productive work on these problems in the future by enabling that work to draw upon connections of the kind made explicit here. That work strikes me as important, as I take it that the thought that our various beliefs or the various propositions to which we assent might be shown to be ultimately arbitrary in terms of their justification is a deeply disturbing one.

In **Section 2** I will give an account of the Five Modes and examine the case for isolating a trilemma produced by the application of three modes within the five. In **Section 3** I will examine Sextus' expression of the Two Modes and argue that these express the same general problem-form as the trilemma, now condensed into a dilemma. I should emphasise that a substantial portion of **Sections 2** and **3** will be spent examining interpretative or historical claims to the effect that the presence of the trilemma or dilemma in question in the Five Modes and the Two Modes was augmented in various ways by Pyrrhonian Sceptics in order to address Dogmatic defences of foundationalism. My intention is not so much to challenge

Improvement," *Phronesis* 59, 3 (2014): 269. See K. Janáček, *Studien zu Sextus Empiricus, Diogenes Laertius und zur Pyrrhonischen Skeptizismus* (Berlin: de Gruyter, 2008): 176-77, for a dissenting view, arguably supported by M. Catapano, "The Two Modes of Scepticism and the Aporetic Structure of Foundationalism," *Méthexis* 29, 1 (2017): 114. This issue does not strongly concern me here. If one should agree that the Two Modes represent a development in the presentation of the Five Modes and think it likely that it is therefore a student or a follower of Agrippa who is responsible for this development, then one can take the term "Agrippan" in the title of this essay to refer to problems with their ancestry in the work of Agrippa's school, rather than problems with their ancestry in the work of one philosopher called Agrippa.

these interpretative or historical claims as it is to head off in advance the philosophical claim that might be made on the back them: that without augmentations of this kind the Sceptical arguments of the Agrippan modes lose some of their efficacy.⁶ Finally, in **Section 4**, I will argue that the general problem-form derived from the Agrippan modes is instantiated in a number of distinct problems, including the epistemic regress problem, the problem of the criterion, and suggest that Hegel's "problem of beginning" should be added to this list.

2. The Five Modes and the Epistemic Regress Problem

The Five Modes, as is the case with all of the Sceptical modes, are intended to bring about a suspension of judgement (*epoché*) concerning some claim under investigation.⁷ They achieve this by the generation of equipollent oppositions, whereby both some claim⁸ that P and some opposing claim that $\neg P$ ⁹ are rendered equally convincing, or are shown to be equally well supported, so that it would be illegitimate to assert the truth of either.¹⁰ Both Sextus and Diogenes present the Five Modes in the same order. I will provide a brief survey of them here.

The first, the mode of dispute, applies when "undecidable dissension about the matter proposed has come about" (*PHI*: 165). If the dissension or disagreement is undecidable, at least on the basis of the evidence currently available, it seems that the reasonable thing to do is to suspend judgement accordingly. The second mode, the mode of infinite regression, suggests that when the evidence provided in support of some claim in fact tends towards an infinite regress of supporting evidence, it then offers no real support as "we have no point from which to begin to establish anything" (*PH*: 166). Given that the claim in question on the matter at

⁶ I am not attributing this philosophical claim to defenders of the various interpretative or historical claims.

⁷ Sextus prefaces his discussion of the modes by saying, "I shall set down the modes through which we conclude with suspension of judgement" (*PHI*: 35).

⁸ As I am introducing Agrippan problems by way of a discussion of Pyrrhonism, I shall tend to use "claim" as a neutral term rather than "proposition" or "belief" as it seems to me to be anachronistic to import a contemporary distinction between propositional and doxastic justification into a reading of Sextus. It seems to me that the Agrippan modes could be applied in both contexts.

⁹ Of course, it need not be P's strict negation, but merely some claim which is incompatible with P and which therefore implies $\neg P$.

¹⁰ I follow K. Vogt, "The Aims of Skeptical Investigation," in *Pyrrhonism in Ancient, Modern, and Contemporary Philosophy*, ed. D. Machuca (Dordrecht: Springer, 2011): 40-41, in taking Sextus' concern to be the rational status of the claim, rather than concerns of a psychological nature about the context in which an individual is likely to be persuaded by the claim.

hand is effectively unsupported, an opposing claim can be asserted as equally convincing, so the reasonable thing to do is again to suspend judgement on the matter.

The third mode, that of relativity, at least as it is described by Sextus, appeals to the relativity of varied perceptions and thoughts concerning the matter at hand to motivate a suspension of judgement over the veracity of any particular perspective (*PH I*: 167). A natural way to read Sextus' account of this mode is as suggesting that, given that some object can appear one way to Perceiver A and another way to Perceiver B, there is no way to know how the object really is, and that we must therefore suspend judgement on the matter.¹¹ The fourth mode is the mode of hypothesis, which applies in those instances where some Dogmatic interlocutor asserts the truth of something "which they do not establish but claim to assume simply and without proof" (*PH I*: 168). This mode points out that the assertion in question appears completely arbitrary since no attempt has been made to support it. It is therefore no more convincing than an opposing assertion that might be made equally arbitrarily, so it is reasonable to suspend judgement on the matter.

Finally, Sextus describes the mode of reciprocity, which covers the familiar, problematic scenario in which the evidence upon which some claim relies for its support includes, at some point, that claim itself, so that its justification is viciously circular (*PH I*: 169). Like the mode of infinite regression, this is taken to undercut the value of that evidence, so that the claim in question is effectively unsupported. An opposing claim can again therefore be equally convincingly asserted, so that the reasonable thing to do is to suspend judgement on the matter. This is only a rough account of the Five Modes, and it obscures a number of important interpretative

¹¹ On Fogelin's reading, addressed below, the mode of relativity appears to have precisely the same function as that of dispute. Hankinson agrees, suggesting that "Relativity, at least in its Agrippan context, should perhaps not be treated as a separate Mode at all" (Hankinson, *The Sceptics*, 185). Sienkewicz, rightly it seems to me, suggests that either relativity reduces to the mode of dispute, or it in fact renders dispute impossible. This would be the case if, rather than the mode of relativity indicating the possibility of dispute by calling attention that the same object can appear one way to Perceiver A and another way to Perceiver B, it instead amounted to the claim that the same object *is* one way, in accordance with the relation it bears to Perceiver A, and *is* another way, in accordance with the relation it bears to Perceiver B (S. Sienkewicz, *Five Modes of Scepticism: Sextus Empiricus and the Agrippan Modes* (Oxford: Oxford University Press, 2019): 147-53). It can look, therefore, as though we are better off ignoring the mode of relativity when paying attention to the Five Modes. Below, however, I will address an alternative reading of the mode of relativity which assigns it a distinctive, important role of its own.

issues, but it will suffice as a starting point for the remainder of this investigation.¹² The Sceptics famously claimed that, rather than the use of the various modes prompting in them a deep unease at the thought that they had no good reason to believe anything, they instead found that a sense of tranquillity (*ataraxia*) followed their suspension of belief (*PHI*: 25-29). This extraordinary suggestion, however, is not under investigation here.

Readers familiar with the *Posterior Analytics* have tended to notice that the combination of the modes of hypothesis, reciprocity, and infinite regress coincides with Aristotle's account of the challenge to the possibility of knowledge through demonstration.¹³ As Aristotle's account suggests, it looks as though there is a specific sceptical problem which can be generated by the application of these three modes together, one which can be used to ask of any claim P whether it has been merely arbitrarily asserted, corresponding to the mode of hypothesis, or whether it depends for its support on some other claim E1. If the latter, then the focus then switches to E1. If this latter claim has been asserted arbitrarily, then this presumably nullifies the support which it offered to P. If, on the other hand, E1 is supported by some claim E2, then the investigation continues. Should this chain of supporting claims eventually repeat, so that E2 turned out to rely upon E1, or indeed P, for its support, then the support for P is thought to be undermined by the circularity of the argument. This corresponds to the mode of reciprocity. If the chain neither terminates somewhere arbitrary nor repeats, then the support is thought to be undermined by the unacceptability of an infinite regress. Thus, together, these three modes look as though they pose a sceptical challenge to the justification of any claim whatsoever.

¹² Barnes, *The Toils of Scepticism*, and Sienkewicz, *Five Modes of Scepticism*, are extended studies of the Five Modes. My brief summary here has the Sceptical modes, for the most part, motivating a suspension of judgement on the grounds that it is the epistemically responsible or reasonable thing to do under the circumstances. Sienkewicz, however, makes a strong case for the claim that the Pyrrhonian Sceptic as presented by Sextus is not entitled to a theoretical framework that allows her to suspend judgement on the basis of what the rational thing to do is (See Sienkewicz, *Five Modes of Scepticism*: 22-46 for a discussion of this issue in the context of the mode of disagreement, for example). I cannot adjudicate this matter here, but see C. Perin, *The Demands of Reason: An Essay on Pyrrhonian Scepticism* (Oxford: Oxford University Press, 2010): 33-58 for an argument that the Sceptic *is* committed to standards of rationality.

¹³ Aristotle, *Posterior Analytics*: 72b, 5-24. I have no intention here of attempting to provide an interpretation of the nature of Aristotle's defence of the possibility of immediate, undemonstrated knowledge.

This combined or systematic use of three of the Five Modes as they appear in the *Outlines* is attributed to Sextus in Barnes' study of the Agrippan modes,¹⁴ and it clearly corresponds to what is today discussed either as the 'epistemic regress problem,'¹⁵ or, acknowledging its ancestry in the presentation of the Five Modes, the 'Agrippan trilemma.'¹⁶ A survey of all of the contexts in which this problem has arisen throughout the history of western philosophy cannot be accomplished here. There is good reason to assign it a fundamental role in the development of German Idealism,¹⁷ and it continues to receive significant attention in contemporary epistemology, where no clear consensus has formed concerning its adequate solution.¹⁸ Although I will suggest later on that the general form of the trilemma occurs in a number of distinct problems, the text of *PHI*: 164-69 reads as though it is problematising the providing of evidence in support of claims in the manner which now occurs in discussions of the epistemic regress problem, and so it is in

¹⁴ See Barnes, *The Toils of Scepticism*, 118. However, see Bett, *How to be a Pyrrhonist*, 123-24 and Sienkewicz, *Five Modes of Scepticism*, 189-91, for suggestions that the Pyrrhonian Sceptic, as described by Sextus, is not entitled to (and largely does not) invoke the three modes in this systematic fashion, but only in a piecemeal fashion. It is possible, as Bett acknowledges, that Agrippa might have intended his modes to have a systematic function, even if elements of Sextus' characterisation of Pyrrhonian scepticism prevent him from endorsing this.

¹⁵ See, for example, A. Cling, "The Epistemic Regress Problem," *Philosophical Studies* 140, 3 (2008): 401-421; T. Kajamies, "A Quintet, a Quartet, a Trio, a Duo? The Epistemic Regress Problem, Evidential Support, and Skepticism," *Philosophia* 37, 3 (2009): 523-34; S. F. Aikin, *Epistemology and the Regress Problem* (New York, NY: Routledge, 2011), 8-43.

¹⁶ See, for example, R. Fogelin, *Pyrrhonian Reflections on Knowledge and Justification* (Oxford: Oxford University Press, 1994), 114, or M. Williams, *Groundless Belief: An Essay on the Possibility of Epistemology* (Princeton, NJ: Princeton University Press, 1999), 183. Both Fogelin and Williams suggest that their choice of title reflects an emphasis that the problem is a trilemma between three equally problematic justificatory alternatives, suspecting that referring to the problem as the epistemic regress problem involves a misleading emphasis upon only one of these. I am not particularly worried about this and will continue to talk about the epistemic regress problem, especially in **IV**. This is because, as I shall explain there, I think that the term "Agrippan trilemma" can also be misleading in this context.

¹⁷ See, for example, R-P. Horstmann, "The Early Philosophy of Fichte and Schelling," in *The Cambridge Companion to German Idealism*, ed. K. Ameriks (Cambridge: Cambridge University Press, 2000), 120-21; T. Pinkard, *German Philosophy 1760-1860: The Legacy of Idealism* (Cambridge: Cambridge University Press, 2002), 94, 98-100; P. Franks, *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism* (Cambridge, MA: Harvard University Press, 2005), 8-10.

¹⁸ Klein suggests that, "Many contemporary epistemologists take the epistemic regress problem as *a*, if not *the* central problem, in epistemology" (P. Klein, "Human Knowledge and the Infinite Progress of Reasoning," *Philosophical Studies* 134, 1 (2007): 1).

the context of this problem that I will continue to discuss the Agrippan modes for the time being.

Given that engagements with the systematic combination of three of the Five Modes both predates and postdates Sextus' account, it is not surprising that there has been a tendency on behalf of philosophers to separate the modes of the trilemma from the modes of dispute and of relativity when engaging with the Five Modes. Thus, for example, Fogelin distinguishes these two from those making up the trilemma by referring to the former as "challenging modes" and the latter as "dialectical modes."¹⁹ The thought governing this division is that the challenging modes, rather than themselves being sufficient to motivate a suspension of judgement, call attention to an actual or even merely possible difference of opinion on whether or not it is the case that P. In doing so, they give rise to an investigation into the evidence that might be put forward in support of holding that P. This investigation is accordingly referred to the dialectical modes, which, at least according to Fogelin, is bound to result in a suspension of judgement, since he holds that "no justificatory program seems to show any prospect of solving the Agrippa problem."²⁰

This way of isolating the trilemma within the context of Sextus' account seems to me both to have been the most influential and to be the most elegant way of recombining the material of the Five Modes. It is the problems that appeal to the modes of the trilemma which will be my focus in discussing the nature of Agrippan problems later on. It is worth acknowledging, however, that there is no clear indication in Sextus' text that the Five Modes are to be divided in this fashion, and when Sextus does suggest that the modes might be used collectively, he mentions all five.²¹ A longer discussion of Sextus' attitude concerning the combination of the three of the Five Modes into a trilemma falls beyond my concerns here.²² It is enough to have shown that the possibility of extracting this trilemma from the Five Modes exists, and that it is this problem which has continued to draw significant philosophical interest from philosophers since Sextus' time.

¹⁹ Fogelin, *Pyrrhonian Reflections on Knowledge and Justification*, 116

²⁰ Fogelin, *Pyrrhonian Reflections on Knowledge and Justification*, 193

²¹ See *PH* I: 170-77. This passage is held to be difficult to interpret, and the tendency has been to reject it as incoherent (See, for example, Barnes, *The Toils of Scepticism*, 114). A sophisticated reading is provided by Sienkewicz (Sienkewicz, *Five Modes of Scepticism*, 157-66), although even this discusses the interrelations only of four of the Five Modes, with the author finding the mode of relativity to be incompatible with the others.

²² Famously, Sextus' use of the Agrippan modes in *PH* II: 20 suggests the combined application of the modes of the trilemma.

Before moving on, however, I should acknowledge that it has recently been suggested by Brennan and Lee that, although we should retain the reading of dispute as a challenging mode, the Agrippan trilemma is better understood as an Agrippan tetralemma because the mode of relativity in fact targets putative self-supporting claims, as distinct from arbitrary assertions and claims which rely upon circular or infinitely regressive chains of supporting claims.²³

This looks as though it is an important point to address. Arguably the most common response to the epistemic regress problem has been to defend some variant of foundationalism, often the suggestion that there are some basic claims which are self-supporting.²⁴ Defenders of this perspective will not be moved by the suggestion that this opens them to the mode of hypothesis, since, they will claim, there is a world of difference between a merely arbitrary assertion, opposition to which can be immediately asserted just as convincingly, and the self-supporting claims which they are endorsing, opposition to which, they suppose, is unreasonable. It looks as though the Sceptic needs a strategy to respond to such a claim, and according to Brennan and Lee, this was the function of the mode of relativity.²⁵

The suggestion made here is that Sextus has mixed up his account of the mode of relativity in the Five Modes with the account of relativity which belongs to the earlier, Aenisideman Ten Modes.²⁶ Instead, Brennan and Lee direct us towards the other classical statement of the Five Modes, that of Diogenes Laërtius. Diogenes describes the mode of relativity quite differently: 'The mode concerned with relativity says that a thing can never be understood in and of itself, but only in relation to something else.'²⁷

²³ Brennan and Lee, "A Relative Improvement." This view receives support in Bett, *How to be a Pyrrhonist*, 119.

²⁴ As Catapano points out, in the historical context of the Pyrrhonian Sceptics, the Stoics would argue that claims made on the basis of certain 'cataleptic' impressions required no support from additional claims (Catapano, "The Two Modes of Scepticism and the Aporetic Structure of Foundationalism," 110, n.12). For a more recent defence of the idea that there can be certain basic beliefs which require no further support, see L. BonJour, "Foundationalism and the External World," *Noûs* 33, s13 (1999): 229-49.

²⁵ Hankinson, by contrast, entertains the notion that self-supporting claims might be considered a limit case for the application of the mode of reciprocity (Hankinson, *The Sceptics*, 189). Brennan and Lee reasonably point out that the mode of reciprocity does not appear to have been used in this way (Brennan and Lee, "A Relative Improvement," 256).

²⁶ The Ten Modes are not my concern here, but their general characteristic is relativity in the sense suggested by Sextus' account of relativity in the Five Modes. Compare *PHI*: 167 with *PHI*: 36-39.

²⁷ Diogenes Laërtius, *The Lives of the Eminent Philosophers*, IX: 89.

Brennan and Lee accordingly suggest that this can be reasonably read not to refer to relativity as described above, but to express a rejection of the idea that one might know that P by considering P in and of itself, and not by appeal to the justificatory relations that P bears to other claims. Their suggestion for the use of the mode of relativity is as follows:

If a Dogmatist claims that something can be known through itself—that an axiom is self-justifying, self-explanatory, self-evident, etc.—then the sceptic who employs the mode of Agrippan Relativity will counter that nothing can be known in this way. Whatever is epistemically grounded, must be grounded in something distinct from itself: thus, anything that is alleged to be grasped through itself is in fact unknown.²⁸

On this reading, the mode of relativity does seem to be directed precisely at the claim that something can be asserted in isolation without further support and yet still be considered to be possessed of some kind of justificatory force, or not to be merely arbitrary. Accordingly, Brennan and Lee suggest that we really ought to refer to the “Agrippan Tetralemma” when considering the traditional horns of the epistemic regress problem. Still, putting aside the historical question of Agrippa’s original statement of the Five Modes, I think that there may be good reasons to continue to operate simply with the trilemma in the context of this essay. The first of these is that the mode of relativity, so understood, seems as though it commits the Sceptic dogmatically to the rejection of the possibility of self-supporting claims. Complex scholarly discussions about the extent to which the Pyrrhonian Sceptic is entitled to any theoretically developed beliefs aside,²⁹ it does not look as though a clear argument is provided in the account of the mode of relativity which would license the Sceptic’s dismissal of some putatively self-supporting claim, still less convince the Dogmatist convinced of its validity to suspend judgement over it.

One might think that a reasonable response here would be to say something similar about the three modes making up the classical trilemma.³⁰ For example, it

²⁸ Brennan and Lee, “A Relative Improvement,” 254.

²⁹ The debate about the extent of the Pyrrhonian Sceptic’s legitimate theoretical commitments is not the focus of this essay, but even according to the defenders of a more moderate interpretation of Sextus on this score (M. Frede, “The Sceptic’s Beliefs,” in *The Original Sceptics: A Controversy*, eds. M. Burnyeat and M. Frede (Indianapolis, IN: Hackett, 1997), 1–24), is the classic statement of this position), the Sceptic is not permitted to hold the kind of philosophically complex beliefs involved in the rejection of the possibility of self-supporting claims.

³⁰ This is not one of Brennan and Lee’s responses. To discuss these would require going deeper into discussions of the interpretation of Sextus than I can afford to here. My goal here is not to dispute Brennan and Lee’s historical conclusions about the Agrippan modes, but merely to head off the suggestion that might be made on the basis of those conclusions, that something like the

might be suggested that it would be equally dogmatic for the Sceptic to simply rule out the possibility of circular justification. I am not sure that the comparison is quite as clear as this, however. The other three modes, as I read them, are designed to make it clear that, for any claim that P , $\neg P$ can be asserted equally convincingly. They achieve this by showing that the claim in question is not distinguishable from an arbitrary assertion. In the case of hypothesis, this involves no more than just pointing out that an opposing claim can be equally as groundlessly put forward. In the case of reciprocity, this is achieved by showing that, despite the appearance of some support for P , this support in fact reduces to simply asserting P , so that again, $\neg P$ can be equally convincingly asserted. In the case of the infinite regress, this is achieved by showing that the apparent support on offer is continuously postponed, so that it never becomes clear that the Dogmatist in question is doing more than arbitrarily asserting P . Showing that this is the case does not involve a dogmatic statement of the illegitimacy of circular or infinite support.

With Brennan and Lee's account of relativity, the case is less straightforward. The suggestion would have to be that the mode of relativity makes it clear that the putatively self-supporting claim is in fact indistinguishable from an arbitrary assertion, but it is not at all obvious how this would be achieved by means of this mode. The statement of the mode suggests no procedure for showing that this is the case, as the modes of the trilemma do. It appears simply to rule out the possibility of self-supporting claims as a matter of principle. It therefore seems to require, at the very least, *more* in the way of dogmatic commitment than the modes making up the trilemma.³¹

Additionally, I would suggest as a second reason for thinking that we can proceed with the traditional extraction of a trilemma from the Five Modes, that the trilemma may very well be all that the Sceptic needs in order to respond to claims that are put forward as self-supporting.³² All that is required is to point out that the claim that a certain claim that P is self-supporting prompts a demand for the justification of this claim in turn. If no evidence for thinking that P is self-supporting can be provided, then it seems indistinguishable from a case open to the mode of hypothesis. If the evidence provided turns on asserting the truth of P ,

relativity mode is essential to the horns of the epistemic regress problem.

³¹ There might be an exception to this case if the original Pyrrhonian Sceptics deploying the Five Modes were engaging with opponents who were defending various claims on the basis of the claim that infinite regresses, for example, can provide genuine support, but I take it that they were not.

³² See Brennan and Lee, "A relative improvement," 270, n.23. Their remarks here suggest that they might accept this point, although the possibility that they are entertaining is not the same as the one I have argued for here.

then the mode of reciprocity applies and P again appears arbitrary. Alternatively, the evidence provided may tend towards an infinite regress, in which case again it may seem as though no support for the claim that P is self-supporting has really been provided, in which case again it appears arbitrary.³³

It therefore seems that the traditional account of the trilemma is sufficient for the Sceptic to suggest that one should suspend judgement in the face of a putatively self-supporting claim. Since that case was the principal reason here for engaging with these concerns about the mode of relativity, I will continue to pay attention primarily to the modes of hypothesis, reciprocity, and infinite regress as I continue to consider the Two Modes, and engage no further with the discussion of the Agrippan tetralemma.

3. The Two Modes

Sextus describes the Two Modes in the following manner:

Since everything apprehended is thought to be apprehended either by means of itself or by means of something else, [the Two Modes] are thought to induce puzzlement about everything by suggesting that nothing is apprehended either by means of itself or by means of something else.

That nothing is apprehended by means of itself is... clear from the dispute which has occurred among natural scientists over, I suppose, all perceivable things and intelligible things – a dispute which is undecidable, since we cannot use either something perceivable or something intelligible as a criterion because anything we may take has been disputed and so is unconvincing.

And for the following reason they do not concede either that anything can be apprehended by means of something else. If that by means of which something is apprehended by means of something else, they throw you into the reciprocal or infinite mode; and if you should want to assume that that by means of which another thing is apprehended is itself apprehended by means of itself, then this is countered by the fact that, for the above reasons, nothing is apprehended by means of itself. (*PHI*: 178-79)

As I suggested earlier, I think that this can be read as a restatement of the argument of the trilemma located in the Five Modes.³⁴ The simplest way to express

³³ See Klein, "Human Knowledge and the Infinite Progress of Reasoning," 14-15, for an extremely clear contemporary use of this strategy. It is worth noting that Klein is explicitly targeting foundationalism and explicitly making use of the modes of the trilemma to do so.

³⁴ Janáček goes further, suggesting, that the presentation of the Two Modes provides the first genuinely systematic account of the use of the Agrippan modes: "I believe that here is where we first find the model according to which the five loosely bound modes receive a logical order" (Janáček, *Studien zu Sextus Empiricus, Diogenes Laertius und zur Pyrrhonischen Skeptizismus*:

this is to suggest that the rejection of the possibility of apprehending something immediately, or by means of itself, corresponds to the application of the mode of hypothesis, while the rejection of the possibility of apprehending something mediately, or by means of something else, corresponds to the application of the modes of reciprocity and infinite regress. Just as the argument of the trilemma appeared to pose a problem for the justification of any claim whatsoever, here the Two Modes are held to “induce puzzlement about *everything*.” However, there are two idiosyncrasies of the presentation of the Two Modes that must be addressed, although I shall attempt to do so as briefly as possible. The first of these is the apparent absence of the mode of hypothesis from the account of the Two Modes and the second is the appeal to the problem of the criterion in the first of the Two. Happily, they can be dealt with together. My contention in both cases is that these idiosyncrasies provide, as was the case with the alternative reading of the mode of relativity examined in the previous section, interesting suggestions concerning the historical use made of the Agrippan modes by the original Pyrrhonian Sceptics, but do not provide philosophical restrictions on later uses of these modes which focus their attention upon the trilemma.

As stated, the first idiosyncrasy is that there is no explicit mention of the mode of hypothesis in the statement of the Two Modes. Instead, it looks as though Sextus references the mode of dispute when rejecting the possibility of unsupported knowledge. This might suggest that if there is a trilemma present in the Two Modes, it consists of dispute, reciprocity and infinite regress.³⁵ Barnes considers this a weakness in the presentation of the Two Modes, complaining that “they ignore the hypothetical mode – which... is a mode of the first importance to the Pyrrhonists.”³⁶ Of course, I agree with Barnes that the mode of hypothesis is of the first importance. As I have sketched the epistemic regress problem in **Section 2**, the Agrippan modes function by drawing attention to the justificatory arbitrariness of claims, and it is the mode of hypothesis that most directly emphasises this. The modes of reciprocity and infinite regress, as I have sketched them, function by stripping away the illusion of support, and rendering the claim in question indistinguishable from an arbitrary hypothesis. In this sense, one might say, hypothesis is the most fundamental of the modes.

176). Translations from Janáček are my own.

³⁵ This is how it has commonly been taken. See, for example, Barnes, *The Toils of Scepticism*, 117-18, P. Woodruff, “The Pyrrhonian Modes,” in *The Cambridge Companion to Ancient Scepticism*, ed. R. Bett (Cambridge: Cambridge University Press, 2010), 226, or Sienkewicz, *Five Modes of Scepticism*, 166.

³⁶ Barnes, *The Toils of Scepticism*, 119.

The tendency among interpreters of the Two Modes who have not been content to regard them as a flawed presentation of the Sceptic's arguments has been to connect Sextus' use of dispute here to the matter of claims that are put forward as foundational or self-supporting, the same issue that concerned us in the discussion of the alternative account of the mode of relativity previously. Specifically, the dominant interpretation of this passage is that while the modes of reciprocity and infinite regression function, in the second of the Two Modes, just as they do in the formula of the epistemic regress problem which we extract from the Five Modes, the function of dispute as it occurs in the first of the Two Modes is to introduce another, distinct problem: the problem of the criterion.³⁷ This is the second idiosyncrasy.

The problem of the criterion is distinct from the epistemic regress problem. The latter targets the attempt to provide evidence in support of a claim, while the former targets the suggestion that a claim can be recognised as true because it is authorised by a criterion of truth.³⁸ Although they are different problems, however, they both make use of the three modes of the trilemma. Accordingly, I shall go on to suggest that both are instances of what I am calling Agrippan problems. For now though, I will provide a brief explanation of the problem of the criterion and explain why it is thought to be invoked in the context of the Two Modes.

Among the Sceptics' Dogmatic opponents, appealing to a criterion in order to argue that a particular claim was true without needing to support it with further evidence was a common strategy.³⁹ Whether the criterion in question is formulated in terms of the human subject making the claim, in terms of the relevant cognitive faculty, in terms of the nature of an appearance which gives rise to a claim, or in terms of some other possibility, the general character of this procedure involves appealing to a criterion where that criterion is a principle which can be used to distinguish truth from falsity. A good criterion, accordingly, would seem to be one that accurately sorts true claims from false ones.

The problem of the criterion works by questioning whether or not the criterion in question is in fact a good one, or why we should suppose that it accurately sorts true claims from false ones. As Sextus would put it, it prompts a

³⁷ Various versions of this interpretation are defended in Hankinson, *The Sceptics*, 189-191; Catapano, "The Two Modes of Scepticism and the Aporetic Structure of Foundationalism;" Sienkewicz, *Five Modes of Scepticism*, 167-77.

³⁸ Here I am following A. Cling, "Reasons, Regresses, and Tragedy: The Epistemic Regress Problem and the Problem of the Criterion," *American Philosophical Quarterly* 46, 4 (2009): 333.

³⁹ Sextus addresses various Dogmatic accounts of purported criteria of truth in *PH* II: 48-79.

dispute about the criterion. It appears that the way in which one would justify a particular criterion is to refer to claims which are known to be true, and which the criterion accurately sorts. These claims, of course, cannot be the same as the original claim, on pain of circularity or reciprocity. But if the Sceptic asks how it is that these other claims are known to be true, then the Dogmatist cannot appeal again to the same criterion, again on pain of circularity. Should the Dogmatist simply assert the validity of a particular claim *or* a particular criterion without being able to authorise or evidence it, then it is no more than an arbitrary hypothesis. And if the Dogmatist is forced to cycle through a non-repeating, never-ending sequence of claims and criteria, this is suggestive of an infinite regress. Sextus provides a brief statement of the problem of the criterion, explicitly referring it, as I have here, to the modes of the trilemma, in *PHI*II: 20.⁴⁰

That Sextus is referring to his treatment of the problem of the criterion in the account of the Two Modes is clear from his account of the first of those modes. There, as we have already seen, he suggests that a dispute about something, be it a matter of perception or of thought, that is held to be apprehended by means of itself will be undecidable, “since we cannot use either something perceivable or something intelligible as a criterion because anything we may take has been disputed and so is unconvincing” (*PHI*I: 178).

Clearly, the text here suggests that the challenge from the Sceptic to the Dogmatist who holds that a certain claim can be apprehended by means of itself in virtue of the fact that it is authorised by what they suppose to be the criterion of truth is to inquire into the reasons for accepting that criterion. The argument will be that the case for the criterion will result in an undecidable dispute, because the Sceptic will refer the case to the modes of the trilemma in exactly the manner sketched above. If the dispute surrounding the criterion is undecidable, we have no compelling reason to accept that the claim in question is in fact a genuinely self-supporting one. This account, coupled with the more straightforward applications of the modes of reciprocity and infinite regress in the case of things which are held to be apprehended by means of something else, constitutes the dominant interpretation of the manner in which the Two Modes are held “to induce

⁴⁰ “In order for the dispute that has arisen about criteria to be decided, we must possess an agreed criterion through which we can judge it; and in order for us to possess an agreed criterion, the dispute about criteria must already have been decided. Thus the argument falls into the reciprocal mode and the discovery of a criterion is blocked – for we do not allow them to assume a criterion by hypothesis, and if they want to judge the criterion by a[nother] criterion we throw them into an infinite regress.

Again, since a proof needs a criterion which has been proved and a criterion needs a proof which has been judged, they are thrown into the reciprocal mode” (*PHI*II: 20).

puzzlement about everything,” or to prompt one to a suspension of judgement over any given claim. It seems to me to make good sense of the text of the *Outlines*.

It is worth emphasising that, in the application of the problem of the criterion to what are put forward as self-supporting claims, it becomes clear that the mode of hypothesis is not in fact absent from the system of the Two Modes in Sextus, since the text of *PHII*: 20 explicitly makes use of it.⁴¹

In addition, the suggestion might be made that the mode of hypothesis is also more directly implicit in the statement of the first of the Two Modes, because, if the claim in question were indeed merely an arbitrary assertion, it would be obviously equipollent to any incompatible alternative.⁴² The only kind of unsupported claim that therefore needs dealing with is the kind which is held to have a special, foundational status which distinguishes it from a merely arbitrary assertion, and so it is this kind of claim which receives focus in the statement of the mode. This strikes me as a plausible reading, but even if it were to be rejected, the mode of hypothesis is still present *a)* in that, at least according to the interpretation I offered earlier, the modes of reciprocity and infinite regress operate by showing that apparently supported claims are in fact indistinguishable from arbitrary hypotheses, and *b)* in the invocation of the problem of the criterion in the first of the Two Modes as we have just seen. It would be a mistake, therefore, to suppose that the mode of hypothesis is absent from the Two Modes.

It seems clear then that the Sceptics making use of the Two Modes intended them to target both claims put forward on the basis of other claims, and those put forward as self-supporting, and thereby to cause trouble for justification in general, in a manner which we would now categorise as an application of the epistemic regress problem. It also seems, on the basis of the text of Sextus' *Outlines*, that their preferred methodology for targeting putatively self-supporting claims was to invoke a different problem, the problem of the criterion, in order to supplement their application of the epistemic regress problem. I do not intend to raise any criticisms about this procedure here.⁴³ However, if it were to be suggested, on the

⁴¹ This point is also common to the accounts of the Two Modes which I have labelled the dominant interpretation. See Hankinson, *The Sceptics*, 191, Catapano, “The Two Modes of Scepticism and the Aporetic Structure of Foundationalism,” 117, and Sienkewicz, *Five Modes of Scepticism*, 171-72.

⁴² See Hankinson, *The Sceptics*, 189-90, for a suggestion to this effect.

⁴³ Indeed, this account seems clearly preferable to the alternative strategy considered in **Section 2**, where the mode of relativity was required in order for the Sceptic to answer the defender of foundationalism, since the appeal to the problem of the criterion offers a clear argumentative procedure for coming to a suspension of judgement over claims which are put forward as having foundational status, while the alternative reading of the mode of relativity appeared simply to

basis of the account of the Two Modes, that a supplementary appeal to the problem of the criterion is required in order for Sceptic applying the epistemic regress problem to arrive at a suspension of judgement in the face of a claim which is held to be self-supporting, I think that we should disagree.

I would instead like to suggest a simpler reading of the Two Modes: one that need not commit the Sceptic simultaneously to appeal to two different problems in order to ensure the strength of her procedure. The suggestion is that the Two Modes can be thought of as a straightforward restatement of the modes of the trilemma, but this time as a dilemma.⁴⁴ Here the reference to undecidable dispute in the account of the first of the Two Modes indicates that undecidable dispute is always possible in the case of something that has been put forward without any support, or arbitrarily. Therefore, the first mode, which states that nothing can be apprehended by means of itself, really expresses the mode of hypothesis. The second mode, which states that nothing can be apprehended by means of something else, problematises any support that might be provided, by pushing that support either towards reciprocity *or* towards an infinite regress. The Two Modes therefore express a dilemma between the arbitrariness of an unsupported claim and the arbitrariness of fundamentally inadequate attempts at support.

This account is in fact suggested by Janáček's reading of *PH* II: 85,⁴⁵ which, as he notices, is an instance of the application of the Two Modes to Dogmatic disputes concerning the existence of truths. As Janáček has it, "The first mode occurs when the Dogmatists arguing with one another claim something without proof. This dispute is undecidable. The second mode occurs when one wants to prove something. The result is either circularity or regress."⁴⁶ This simplified application of the Two Modes does seem amenable to the idea that the dispute here

reject them as a matter of principle.

⁴⁴ To this extent, the account of the Two Modes might be said to anticipate Kajamies' claim that the epistemic regress problem can be expressed, at its simplest, as a "duo" (Kajamies, "A Quintet, A Quartet, A Trio, A Duo?," 533-34).

⁴⁵ "There is a dispute about truths among the Dogmatists; for some say that some things are true, and some that nothing is true. And it is not possible to decide the dispute; for if you say that some things are true, you will not be found convincing if you say it without proof, because of the dispute; and if you actually want to bring a proof... if you say that the proof is true you fall into the reciprocal argument and in addition you will be asked for a proof of the fact that it is true – and another proof for that, and so on *ad infinitum*" (*PH* II: 85).

⁴⁶ Janáček, *Studien zu Sextus Empiricus, Diogenes Laertius und zur Pyrrhonischen Skeptizismus*, 180

is a symptom of the presence of arbitrary hypothesis. It also avoids necessitating an additional discussion of the problem of the criterion in the case of the first mode.⁴⁷

At this point, however, one might wonder how this account addresses the matter that the appeal to the problem of the criterion was thought to address in the dominant interpretation of the Two Modes discussed above. The objection might be that, by restricting the reading of the Two Modes to the options of the trilemma and shifting focus from the manner in which the mode of dispute confronts the advocate of a foundationalist response to the epistemic regress problem with the problem of the criterion, I have robbed the Two Modes of some of their argumentative efficacy. I would respond in exactly the same manner as I responded to the case for insisting that the mode of relativity must be added to the trilemma in **Section 2**. In the face of what is presented as a basic or foundational claim, although the Sceptic certainly might apply the problem of the criterion if the Dogmatist appeals to a criterion to authorise the claim, she could just as well simply ask what the evidence is for supposing that this claim is in fact an instance of such a self-supporting, immediately justified claim. Even if an appeal to a criterion is made, the Sceptic can just as well ask what evidence there is for supposing that this claim *is* in fact authorised by that criterion. There is thus no real need for the Sceptic to diverge from the practice of asking for supporting evidence. The epistemic regress problem as presented in this account of the Two Modes can then apply in a manner which is as problematic as ever for the Dogmatist to respond to, without the Sceptic necessarily having to engage in additional discussion of the problem of the criterion.

I conclude that the most fundamental kind of sceptical problem that we inherit from Sextus and the Pyrrhonian Sceptical tradition can be expressed at its simplest in terms of a dilemma, as it is in the case of the Two Modes. On the one hand one confronts the apparent arbitrariness of a claim that is made without any support, and on the other the apparent arbitrariness of a claim whose support can be shown to be fundamentally inadequate. And in fact, in the discussion above, we have already encountered two distinct problems which appeal to this dilemma.⁴⁸ I would now like to suggest that these are not the only two.

⁴⁷ This is not to dispute the dominant interpretation of *PH* I: 178-79 put forward by Hankinson, Catapano, and Sienkewicz, which tie the first of the Two Modes to the problem of the criterion. It is merely to suggest that, at least on one occasion in Sextus' works, he appears to apply the Two Modes in this more straightforward manner, without needing to invoke a second problem.

⁴⁸ The discussion so far has taken place in terms of the epistemic regress problem, but I take it that it is clear that the problem of the criterion can also be expressed in terms of a dilemma: either *a*) a particular claim *or* a particular criterion is asserted merely arbitrarily, or *b*) inadequate attempts are made to authorize claims and criteria, where those attempts result in circular or

4. Agrippan Problems

At this point we have examined both the Five Modes and the Two Modes. I have made a case for isolating the epistemic regress problem as a trilemma of modes from within the former and argued that this problem can also be presented in a more streamlined manner, as a dilemma, by appealing to the latter. I have also argued that although the application of this problem might be augmented by appealing to an alternative account of the mode of relativity, or to the problem of the criterion, the modes of the trilemma alone are enough to pose a serious challenge to justification of any claim. Finally, we have also encountered, however briefly, two distinct problems which make use of the Agrippan modes: the epistemic regress problem and the problem of the criterion. I now wish to examine this further, and to suggest that these two are not the only ‘Agrippan problems,’ where this term refers to a problem which uses the Agrippan modes to argue for a sceptical conclusion.

Before going further, however, I would like to briefly address a matter of terminology. I am by no means the first person to notice that different problems make use of the Agrippan modes. Sankey, for example, suggests that the problem of the criterion is an instance of “a more general form... sometimes known as *Agrippa’s Trilemma*.”⁴⁹ I am not inclined to adopt this usage because, as I have already noted, the same term is often used to refer specifically to the epistemic regress problem.⁵⁰ Cling offers an excellent analysis of the epistemic regress problem and the problem of the criterion and suggests that they are both instances of what he refers to as “the paradox of reasons,”⁵¹ where this term indicates problems concerning justification which force one towards the Agrippan modes which have been our focus so far. In light of their Pyrrhonian heritage, I suggest that we refer to problems of this kind simply as ‘Agrippan problems.’

Cling also, accurately in my view, indicates that the epistemic regress problem and the problem of the criterion do not exhaust the range of Agrippan

infinitely regressive reasoning.

⁴⁹ H. Sankey, “Epistemic Relativism and the Problem of the Criterion,” *Studies in History and Philosophy of Science* 42, 4 (2011): 562.

⁵⁰ One might also be uncomfortable with insisting on the “trilemma” formulation; on the one hand because, as I have suggested in light of my reading of the Two Modes, it might, at its simplest, be expressed as a dilemma, and on the other because the sheer variety of responses to the epistemic regress problem that have now been developed by epistemologists renders the term “trilemma” rather misleading (See Aikin, *Epistemology and the Regress Problem*, 46, for the suggestion that it might in fact be a “*hexalemma*”).

⁵¹ Cling, “Reasons, Regresses, and Tragedy,” 338.

problems. He suggests that such problems might target proposals for “evidence that a proposition is true; a criterion of truth that sanctions a belief; an epistemic principle that a belief would be a case of knowledge or justified belief; an account of how the state of affairs described by a proposition is possible; factors that would make a belief valuable whether or not it is true, likely to be true, or reasonably believed to be true...”⁵² Elsewhere, Franks suggests this kind of problem can also occur in a metaphysical, rather than only an epistemological context, as one concerning the grounds of the reality of things, and attributes engagement with such a problem to Schelling, in particular.⁵³ Agrippan problems, or sceptical arguments to the effect that beliefs that we hold to be justified are in fact quite arbitrary seem to arise in a wide variety of contexts, and across the history of philosophy.

Here I would like to make a small, rather unusual addition to the list of recognised Agrippan problems by suggesting that a distinctive one occurs at the beginning of Hegel’s *Science of Logic*, in which he attempts to provide a rigorous derivation of the fundamental categories of thought, believing that earlier accounts of the categories such as those of Aristotle or Kant had not been properly justified.⁵⁴ There we encounter the following passage:

The beginning of philosophy must be either *something mediated* or *something immediate*, and it is easy to show that it can be neither the one nor the other; so either way of beginning runs into its rebuttal.⁵⁵

The formulation of this problem, which we can call the ‘problem of beginning,’ seems to echo the expression of the Two Modes examined earlier.⁵⁶ The rejection of immediacy suggests the justificatory inadequacy of beginning with an arbitrary hypothesis and the rejection of mediation suggests the justificatory inadequacy of beginning with something that depends on fundamentally

⁵² Cling, “Reasons, Regresses, and Tragedy,” 338.

⁵³ See Franks, *All or Nothing*, 19, n.11.

⁵⁴ See G.W.F. Hegel, *Science of Logic*, trans. G. di Giovanni (Cambridge: Cambridge University Press, 2010), 525.

⁵⁵ Hegel, *Science of Logic*, 45. Translation altered.

⁵⁶ It turns out that I am not alone in recognising this resemblance. The same suggestion is made by Hentrup (M. Hentrup, “Hegel’s Logic as Presuppositionless Science,” *Idealistic Studies* 49, 2 (2019): 151). Hentrup and I have arrived at this conclusion quite independently of one another, however, which I take to speak in favour of its plausibility as an interpretative claim. Hentrup, however, seems to take Hegel to be attempting, in his *Logic*, to solve a particular problem expressed in Sextus’ account of the Two Modes. I am merely reading the latter as expressing a more general problem-form of which the problem that Hegel is attempting to solve is just one instance.

inadequate support. The allusion to a rebuttal or refutation suggests the sceptical strategy of generating an equipollent opposition in order to motivate a suspension of judgement. It is not surprising that Hegel should find that he is confronted with a sceptical problem in this context.⁵⁷ In attempting to provide an account of the fundamental categories of thought, he is often at pains to emphasise that strict standards for the justification of such an account must be met.⁵⁸ Of course, it is precisely attempts at justification that Agrippan problems tend to target. Accordingly, I will suggest that Hegel's problem of beginning amounts to a distinctive Agrippan problem, but first an immediate objection presents itself to this claim: Hegel's statement of the problem appears to recall the following passage from the opening of Fichte's early *Wissenschaftslehre*:

Our task is to *discover* the primordial, absolutely unconditional first principle of human knowledge. This can be neither *proved* nor *defined*, if it is to be an absolutely primary principle.⁵⁹

Here again the rejection of an appeal to a principle which has been merely arbitrarily defined or to one which problematically defers the matter of justificatory authority resembles the presentation of the Two Modes. But in Fichte's case it is clear that this is not a distinctive Agrippan problem, but merely an instance of the application of the epistemic regress problem. Fichte's description of his task in the passage above indicates that he shares a commitment common to much of post-Kantian German Idealism: that of deriving a systematic set of theoretical and practical positions from a single, fundamental or basic principle.⁶⁰ Given this commitment, Fichte's response to a sceptic who challenges any one of his philosophical claims is therefore to claim that they are all ultimately adequately supported by the fundamental principle which he identifies. Although the specifics

⁵⁷ His own engagement with the Pyrrhonian Sceptical tradition has long been acknowledged, and is perhaps best exemplified in Hegel, "On the Relationship of Skepticism to Philosophy." A longer discussion of Hegel's engagement with Pyrrhonism cannot be accomplished here.

⁵⁸ See, for example, his insistence that in providing such an account of the categories of thought, that they "must be exhibited in their *necessity* and it is essential that they be *derived*" (G.W.F. Hegel, *Encyclopaedia of the Philosophical Sciences in Basic Outline: Part I: Science of Logic*, trans. K. Brinkmann and D.O. Dahlstrom (Cambridge: Cambridge University Press, 2010), §42).

⁵⁹ J.G. Fichte, *Science of Knowledge (Wissenschaftslehre)*, trans. P. Heath and J. Lachs (New York, NY: Meredith Corporation, 1970): 93. The resemblance of the passage from Hegel's *Logic* to Fichte's passage is noted by di Giovanni in his translation of the former, cited above.

⁶⁰ This feature of German Idealist philosophy is discussed in E. Förster, *The Twenty-Five Years of Philosophy: A Systematic Reconstruction*, trans. B. Bowman (Cambridge, MA: Harvard University Press, 2012), 155–64; Pinkard, *German Philosophy 1760–1860*, 98–99; and Franks, *All or Nothing*, 386–87 (and throughout that work).

of Fichte's solution to the problem of identifying such a principle do not concern us here, it should be clear that Fichte defends a kind of foundationalist response to the epistemic regress problem.⁶¹

If the passage from Hegel's *Logic* recalls Fichte's problem of identifying a first principle, one might assume that Hegel, too, is engaged in providing a foundationalist response to the epistemic regress problem when he first sets out and then attempts to solve his problem of beginning, in which case there is no distinctive Agrippan problem to be found here. As with Fichte, the details of the solution Hegel provides to his problem need not concern us now,⁶² but I think that there is a good case to be made for thinking that, in Hegel's case, the problem in question is not just another instance of the epistemic regress problem, along with the assumption that some sort of foundationalism must provide a solution.

Put simply, the case turns on two features of Hegel's project. Firstly, he arguably is committed to foundationalist response to the epistemic regress problem at least in the context of justifying the account of the fundamental categories of thought that he develops in his *Logic*,⁶³ and secondly, the solution to his problem of beginning is not that response. We can treat these, briefly, in turn.

Although it is unlikely to be universally accepted, it is at least far from uncommon to attribute to Hegel a version of the same goal as that of Fichte seen above: that of identifying an unconditional principle to which other philosophical claims can appeal for their support. According to such an account, Hegel's proposal for such a fundamental principle is what he refers to as 'the concept.' The details of Hegel's account of the concept do not concern us here.⁶⁴ What matters is that it

⁶¹ An interesting discussion of Fichte's defence of an unconditional first principle as a response to scepticism is provided in D. Breazeale, "Fichte, Skepticism, and the 'Agrippan Trilemma,'" *Fichte-Studien* Band 44 (Leiden: Brill, 2017): 3-16.

⁶² One interpretation of Hegel's solution to his problem of beginning is provided in Hentrup, "Hegel's Logic as Presuppositionless Science," 153-60. I provide a somewhat different interpretation in R. Dunphy, "Hegel and the Problem of Beginning," *Hegel Bulletin* (forthcoming).

⁶³ I provide no discussion here of how Hegel's logical material relates to his discussion of the making of empirical judgements in his *Philosophy of Spirit*, nor of how Hegel understands empirical judgements to be justified. Hegel's problem of beginning, as I understand it, is an Agrippan problem local to the epistemology of logic, as the term is treated in Hegel's work.

⁶⁴ A compelling account of Hegel's notion of the concept which I take to support this view is provided in B. Bowman, *Hegel and the Metaphysics of Absolute Negativity* (Cambridge: Cambridge University Press, 2013), 31-43. Bowman suggests that Hegel's concept is the structure of self-referential, autonomous negation which he takes to function as "the absolute foundation of Hegelian logic and metaphysics" (Bowman, *Hegel and the Metaphysics of Absolute Negativity*, 52).

plays the key role in the kind of foundationalism that Hegel endorses in the context of his *Logic*. He claims, for example, that “the concept is to be regarded indeed, not just as a subjective presupposition but as *absolute foundation*.”⁶⁵ It is back to the concept, as foundational principle, that the account of the various categories explored in Hegel’s *Logic* are traced.⁶⁶

The reason that Hegel’s problem of beginning cannot be, as in the case of Fichte, an attempt to provide an unconditional first principle, is that Hegel’s elaboration of such a principle, the concept, is not provided as the solution to the problem of beginning. The elaboration of the concept does not take place until the third book of Hegel’s *Logic*, while the work *begins* instead with the abstract category of “pure being,” which is emphatically not an unconditional first principle. Indeed, its coherence appears to be quickly rejected in favour of a discussion of more determinate categories.⁶⁷

The crucial point here, which I cannot explore in detail, is that Hegel, unlike Fichte, does not begin with a foundational principle and derive philosophical claims from there, but envisages the setting out of such a foundational principle as the result of a process of philosophical reasoning.⁶⁸ Thus, after suggesting that the *Logic* begin not with its fundamental principle but with the abstract category of pure being, Hegel claims that, in the context of the argument of his *Logic*, “*progression* is a retreat to the *ground*, to the *origin* and the *truth* on which that with which the beginning was made, and from which it is in fact produced, depends.”⁶⁹ This is reiterated in the passage cited above in which Hegel emphasised that his account of the concept is one of an absolute foundation, where he acknowledges that he found it necessary to begin with something abstract, “the foundation of which... must therefore be sought.”⁷⁰ From this we can conclude that Hegel’s problem of beginning is not an expression of the challenge he faces in providing a foundationalist response to the epistemic regress problem in the context of an account of the categories of thought, but instead the problem of where to begin an investigation of the categories of thought in such a way that that the beginning is not problematically arbitrary, even though it does not constitute

⁶⁵ Hegel, *Science of Logic*, 508.

⁶⁶ See, for example, Franks, *All or Nothing*, 377–79, for another argument to this effect.

⁶⁷ Hegel, *Science of Logic*, 59–82. A good discussion of this topic is provided in R. Pippin, *Hegel’s Realm of Shadows: Logic as Metaphysics in The Science of Logic* (Chicago, IL: University of Chicago Press, 2019), 183–210.

⁶⁸ See Bowman, *Hegel and the Metaphysics of Absolute Negativity*, 43–61 and 166–200 for an account of Hegel’s model of philosophical demonstration.

⁶⁹ Hegel, *Science of Logic*, 49.

⁷⁰ Hegel, *Science of Logic*, 508.

any kind of self-supporting foundation. I suggest that this problem constitutes a distinctive, perhaps idiosyncratically Hegelian, Agrippan problem.

A brief examination of Hegel's remarks in the immediate context of the problem provides us with a clearer understanding of what Hegel perceives to be the inadequacy of beginning either with something mediated or with something immediate. He spells this out in the language of presuppositions:

Being is what makes the beginning here; it is presented indeed as originating through mediation... and the presupposition is of a pure knowledge which is the result of finite knowledge, of consciousness. But if no presupposition is to be made, if the beginning is itself to be taken *immediately*, then the only determination of this beginning is that it is to be the beginning of logic, of thought as such.⁷¹

We need not be lured into a discussion of the more obscure elements of this passage, nor of the sense in which Hegel takes the category of being to somehow solve the problem he is considering.⁷² I want only to point out that this passage makes it clear that by "beginning with something mediated" Hegel means beginning with something which relies upon some presuppositions to demonstrate its validity, while by "beginning with something immediate" he means beginning with something that presupposes nothing. I take it that this again shows that Hegel's problem of beginning shares the dilemmatic form of an Agrippan problem, where the dilemma is between the problematic arbitrariness of something asserted without any support, and the problematic arbitrariness of something asserted on the basis of fundamentally inadequate support. In Hegel's case, there is a question mark hanging over the legitimacy of what is presupposed at the beginning of a logical investigation, and therefore on the beginning itself,⁷³ while an attempt to

⁷¹ Hegel, *Science of Logic*, 47-48.

⁷² I provide an account of the way in which Hegel takes the category of being to solve his 'Agrippan' problem of beginning in Dunphy, "Hegel and the Problem of Beginning."

⁷³ Hentrup takes it that what is problematic about relying on presuppositions in this context is that it defers the matter of the justification of the beginning indefinitely, so that it is an appeal to the mode of infinite regression (Hentrup, "Hegel's Logic as Presuppositionless Science," 151). Hegel's remarks to the effect that a science of logic is not entitled to presuppose the findings of other sciences, rely upon axioms or a recognised method, or even assume an account of its own subject matter at the beginning, suggest to me another possibility. It may be that the apparent support offered by what is presupposed at the beginning of such a science is undermined precisely because all of its basic principles, its "forms of reflection... rules and laws of thinking... are part of its content and they first have to be established within it" (Hegel, *Science of Logic*: 23). In other words, Hegel might not need to argue that the presuppositions in question form a vicious circle or tend towards an infinite regress, but instead could argue that they constitute the 'arbitrariness-of-something-based-on-fundamentally-inadequate-support' horn of an Agrippan

begin without presuppositions is simply to begin with something that is obviously arbitrary as it has nothing supporting it. In the latter case, a Sceptic can oppose any suggested beginning to an alternative that has equally groundlessly been put forward. In the former case, because the validity of what has been presupposed has not been established, a Sceptical opponent is again entitled to hold that it can be equally convincingly opposed to an alternative. Either way, it looks as though a sceptical rebuttal faces any attempt to get Hegel's account of the categories of thought off the ground.

Hegel's problem of beginning is therefore an Agrippan problem which challenges the very possibility of beginning an investigation into the fundamental categories of thought in a manner which is not vitiated by the problematic arbitrariness of the starting point of the investigation, just as the epistemic regress problem challenges the very possibility of holding a claim to be supported by evidence without that evidence being revealed to be fundamentally arbitrary, or the problem of the criterion challenges the very possibility of authorising a claim by appealing to a criterion of truth without both claim and criterion appearing to be problematically arbitrary. Problems of this kind, Agrippan problems, have preoccupied philosophers from antiquity, through the period of German Idealism, and up to the present day.⁷⁴ Their pervasiveness suggests a deep unease concerning the extent to which the content of our thought can escape the threat of being revealed as fundamentally arbitrary. If, as I do, one finds the Pyrrhonian Sceptics' claims to discover tranquillity in the eye of this storm implausible, it seems that there is little choice but to continue in the attempt to solve problems of this kind, although it is certainly easier said than done.⁷⁵

problem because they are presupposed *at the beginning* of science that permits no such thing, on his conception of it. I cannot explore this topic further here.

⁷⁴ It should be noted that I am not suggesting here that because there is a wide variety of Agrippan problems, a successful solution to one will share the same form as a successful solution to another. Such a discussion must take place elsewhere.

⁷⁵ This paper was written during a fellowship at the Maimonides Centre for Advanced Studies at the University of Hamburg.

GROUP TESTIMONY: DEFENDING A REDUCTIONIST VIEW

Domingos FARIA

ABSTRACT: Our aim in this paper is to defend the reductionist (or deflationist) view on group testimony from the attacks of divergence arguments. We will begin by presenting how divergence arguments can challenge the reductionist view. However, we will argue that these arguments are not decisive to rule out the reductionist view; for, these arguments have false premises, assuming dubious epistemic principles that testimony cannot generate knowledge and understanding. The final part of this paper will be devoted to presenting the advantages of the reductionist approach to explaining the phenomenon of group testimony.

KEYWORDS: collective epistemology, group epistemology, group testimony, divergence arguments

1. Group Testimony: Who Is the Source of Knowledge?

We gain a lot of knowledge through the testimony of others; many of our beliefs are learned from the spoken or written word of others. The traditional problem in this field has to do with the epistemic status of these beliefs and whether or not a receiver needs positive reasons to accept the testimony of a sender. Following the tradition of David Hume, it is argued that receivers must possess positive reasons in order to be justified in accepting the testimony of senders. In contrast, following the tradition of Thomas Reid, it is stated that though the presence of positive reasons is not necessary to acquire testimonial justification or knowledge, the absence of negative reasons is necessary. Generally, the focus of analysis in these cases is the testimony of individuals.

However, we can also acquire knowledge through the testimony of groups. For example, given the collaborative work done in science, we gain scientific knowledge through the testimony of research groups. Similarly it seems that we can acquire knowledge through the testimony of collective entities such as organizations, companies, clubs, churches, among others. In this paper, we want to deal with new problems on testimony related to collective epistemology: When we acquire knowledge from the testimony of a group, who is the source of this knowledge? Is this source reducible to an individual (or set of individuals) in the group? Or, instead, is this source not reducible to any element of the group, and

thus the group itself is that source of testimonial knowledge? In other words, can groups testify knowledge that their individual members lack?

There are two main views that answer these problems: reductionism (or deflationism) and non-reductionism (or inflationism).¹ On the one hand, the reductionist view holds that a group's testimony that p is reducible to the testimony of at least one individual in that group. In this sense, a group testifying that p means that at least one individual member of the group would testify that p if the relevant opportunity arises.² Here we want to focus on a minimal version of reductionism, which provides only a necessary condition, but not a sufficient condition, for group testimony. More precisely, this minimal view can be formalized as follows:

Reductionism =_{df} Necessarily, a group g testifies that p only if at least one individual i is both a member of g and testifies that p .

$$\Box(Tpg \rightarrow \exists i(i \in g \wedge Tpi))$$

This view needs some qualifications. Namely, the testimony of an individual member i of g must meet certain conditions, such as the following: (1) i is authorized by g to provide the testimony; (2) i provides testimony as a member of g . Thus, if i is not licensed to testify on behalf of a group g , or if i testifies as private individual or member of a distinct group g^* , then her testimony is not considered group g testimony.³ Lackey⁴ argues for a slightly different version of reductionism, according to which group testimony is reducible to the group's spokesperson who does not have to be a member of the group. However, here we will assume that if there is a spokesperson who speaks on behalf of a given group, then that individual is somehow collaborating with the group and, therefore, belongs in a broad sense to that group.⁵ In short, according to reductionism, the

¹ In this context of collective epistemology, the use of the terms "reductionism" and "non-reductionism" cannot be confused with the reductionism of the Hume tradition and the non-reductionism of the Reid tradition in the context of justifying individual testimony.

² See Jennifer Lackey, "A Deflationary Account of Group Testimony," in *Essays in Collective Epistemology*, ed. Jennifer Lackey (Oxford University Press, 2014), 64–94, <https://doi.org/10.1093/acprof:oso/9780199665792.003.0004> and Jesper Kallestrup, "Groups, Trust, and Testimony," in *Trust in Epistemology* (Routledge, 2019), 136–58, <https://doi.org/10.4324/9781351264884-6>.

³ These qualifications are evidenced by Kallestrup (*ibid.*).

⁴ "A Deflationary Account of Group Testimony."

⁵ We are assuming that groups may have different membership levels or status. In a strict sense a member of a group is one who is properly designated as such (for example by leaders or operational members, in an informal or formal context, and according to the rules of that group). In a broader sense, a member of a group is one who is contributing to the functioning of that

epistemic credentials of group testimony are reducible to the epistemic credentials of at least one authorized individual testimony who belongs in a narrow or broad sense to that group.

On the other hand, the non-reductionist view is the negation of reductionism. In this view the group itself is the source of the testimonial knowledge. For, it holds the possibility that the testimony of a group is irreducible to the testimony of all or some of its individual members. So, a group testifying that p cannot be understood in the sense that at least one individual member of the group would testify that p if the relevant opportunity arises. More specifically, we can formalize non-reductionism as follows:

Non-reductionism =_{df} Possibly, a group g testifies that p even when no individual member of g testifies that p .

$$\Diamond(Tpg \wedge \neg \exists i (i \in g \wedge Tpi))$$

Which of these perspectives is the most plausible? In this paper, we want to defend the reductionist view against divergence arguments. We will begin in section 2 by presenting how divergence arguments can challenge the reductionist view. However, in section 3, we will argue that these arguments are not decisive to rule out the reductionist view. Section 4 will be devoted to presenting advantages of the reductionist approach to explaining the phenomenon of group testimony.

2. Divergence Arguments Against Reductionism

So-called *divergence arguments* are the main motivation for non-reductionism. These arguments aim to show that reductionism is false because it is possible for a group to testify that p while none of its individual members testifies p (or is able to testify p). Thus, concerning testimonial knowledge, there is an epistemic divergence between groups and their individual members. In support of this divergence there are two strong counterexamples against the reductionist view. The first counterexample presents a case in which a group clearly testifies

group; in other words, a member is someone who plays some kind of functional role for the group. Following the proposal of David Strohmaier, "Group Membership and Parthood," *Journal of Social Ontology* 4, 2 (2018): 132, <https://doi.org/10.1515/jso-2018-0016>, "a part of a group is a member of this group if, and only if, it is an agent and it is appropriately designated to contribute to the group's functioning. (...) The members are not designated as members but as contributing to the group's functioning." A spokesperson seems to satisfy these conditions to be a member of a group. Namely, the spokesperson is a member of a group because he is an agent and, moreover, has been appropriately designated as contributing to its functioning (by playing the functional role of spokesperson).

knowledge which its individual members lack. This case is inspired by J. Adam Carter:⁶

REPORT CASE: A small committee of art experts has been called in to determine the veracity of a very rare pottery piece. This committee uses the best methods of science to distinguish genuine and fake antiquities. By following strictly professional and scientific norms they come to a conclusion: the dating of this pottery piece is approximately 14,000 BC – we call this proposition q to abbreviate. Suppose this committee writes a public report with that conclusion q . However, each member of this committee is a young earth creationist, and therefore privately each believes that the earth and the very rare pottery piece are less than 6,000 years old.

This REPORT CASE is a case where not a single member of the group in question believes that q and so, each of its individual members fails to know that q , yet recipients or readers of the public report can nonetheless acquire knowledge that q on the basis of the group's testimony that q . But if each one fails to believe and to know the proposition q , then no individual member can be the source of testimonial knowledge that q . This is because if a sender doesn't know that q , then there is no way that a recipient can come to know that q through believing the sender's testimony.⁷ Instead, the group itself is the source of testimonial knowledge, given that the group itself knows that q .⁸ If this case is plausible, reductionism is false.

⁶ "Group Knowledge and Epistemic Defeat," *Ergo, an Open Access Journal of Philosophy* 2, 20190926 (2015), <https://doi.org/10.3998/ergo.12405314.0002.028>.

⁷ This premise is defended by Angus Ross, "Why Do We Believe What We Are Told?" *Ratio*, 1 (1986): 69–88, Michael Dummett, "Testimony and Memory," in *Knowing from Words* (Springer Netherlands, 1994), 251–72, https://doi.org/10.1007/978-94-017-2018-2_12, Robert Audi, "The Place of Testimony in the Fabric of Knowledge and Justification," *American Philosophical Quarterly* 34, 4 (1997): 405–22, <https://doi.org/10.2307/20009910>, Tyler Burge, "Content Preservation," *Philosophical Issues* 6 (1995): 271, <https://doi.org/10.2307/1523046>; "Interlocution, Perception, and Memory," *Philosophical Studies* 86, 1 (1997): 21–47, <https://doi.org/10.1023/a:1004261628340>, Steven L. Reynolds, "Testimony, Knowledge, and Epistemic Goals," *Philosophical Studies* 110, 2 (2002): 139–61, <https://doi.org/10.1023/a:1020254327114>. The main idea of this premise is that the testimony only transmits, but does not generate new knowledge. In other words, testimony only disseminates knowledge, does not discover or generate it.

⁸ Non-reductionist or inflationary accounts of group knowledge support the possibility that knowledge attributions apply to groups while they do not apply to their members. See, for example, Margaret Gilbert, "Collective Epistemology," *Episteme* 1, 2 (2004): 95–107, <https://doi.org/10.3366/epi.2004.1.2.95>, Raimo Tuomela, "Group Knowledge Analyzed," *Episteme* 1, 2 (2004): 109–27, <https://doi.org/10.3366/epi.2004.1.2.109>, Kay Mathiesen, "Can Groups Be Epistemic Agents?" in *Collective Epistemology*, eds. Hans Bernhard Schmid, Daniel Sirtes, and Marcel Weber (Ontos, 2011), 23–44, Alexander Bird, "When Is There a Group That Knows?" in

Another important counterexample is based on *distributed cognition* in which there is a division of cognitive labor within a group. In such a case the task of producing knowledge is divided into subtasks. Each subtask is assigned, depending on the area of expertise, to different individual members. However, no member of the group is able to grasp each other's tasks. Thus it seems that the production of knowledge, and its consequent testimony, is not reducible to individual members, but is something that depends on the group itself. In order to clarify this point, we can imagine a case inspired by Alexander Bird:⁹

TALK CASE: Dr. X is a physicist and Dr. Y is a mathematician. Both are collaborating on a project to demonstrate the truth of the conjecture q , but each one works alone and without communicating. Suppose they agreed in advance with an assistant, who only knows how to apply *modus ponens*, to give a talk showing that q just in case the assistant receives independently from Dr. X the proof that p is true and from Dr. Y the proof that $p \rightarrow q$. Based on empirical experiments Dr. X shows that p ; while based on pure mathematics Dr. Y shows that $p \rightarrow q$. The assistant applies *modus ponens* and gives the talk.

In this case the research team seems to know and to testify that q , and such knowledge is acquired and understood by an audience of scientists to whom the assistant is giving the talk. However, none of the research team members individually knows and is able to testify that q . Neither the assistant himself, who is giving the talk, is the source of the testimonial knowledge that q ; given that he does not have the ability to understand the demonstration that p and the proof that $p \rightarrow q$, as well as he is unable to grasp the meaning of q . He is simply reading the result, not understanding the body of information or domain he is talking about. In this regard, Kallestrup¹⁰ holds that:

The testifier is the group itself. The point here is not so much that the assistant lacks knowledge of q , but that he even lacks the required expertise to grasp such a complex proposition. Because nobody can properly assert a proposition they do not understand, the assistant cannot be regarded as testifying that q . Nor can either Drs. X and Y be said to testify q . Nothing about the way the case is described suggests their linguistic behavior amounts to an assertion that q . Dr. X asserts that p but not $p \rightarrow q$, whereas Dr. Y asserts $p \rightarrow q$ but not p , and so neither

Essays in Collective Epistemology, ed. Jennifer Lackey (Oxford University Press, 2014), 42–63, <https://doi.org/10.1093/acprof:oso/9780199665792.003.0003>, and Deborah Tollefsen, *Groups as Agents* (Polity, 2015).

⁹ "Social Knowing: The Social Sense of 'Scientific Knowledge'," *Philosophical Perspectives* 24, 1 (2010): 23–56, <https://doi.org/10.1111/j.1520-8583.2010.00184.x>; "When Is There a Group That Knows?" 57–58.

¹⁰ "Groups, Trust, and Testimony," 140.

asserts q on the basis of *modus ponens*.

Are these counterexamples decisive in defending non-reductionism and rejecting reductionism? In the next section we will argue that these are not good counterexamples.

3. Defeating Divergence Arguments

In the previous section, we saw two main ways to attack the reductionist view. Now we want to argue that none of them is plausible.

3.1 Problems With the First Argument

Analyzing the first case, REPORT CASE, we can highlight that the central argument can be reconstructed as follows: No single member of the group g knows that q . If no single member of g knows that q , then a recipient R cannot come to know that q on the basis of the testimony that q provided by a single member of g . Thus, R cannot come to know that q on the basis of the testimony that q provided by a single member of g . However, REPORT CASE shows that R can come to know that q on the basis of the testimony that q provided by g itself, given that g itself can know that q . Therefore, if this is so, then the source of testimonial knowledge is g itself, not being reducible to its individual members. Is this a good argument? Note that the argument works only if the following premise is true:

(K) If a sender S doesn't know that p , then a recipient R cannot come to know that p on the basis of the testimony that p provided by S .

Premise (K) underlies the idea that the testimony is merely transmissive and cannot itself generate new knowledge. In other words, the testimony can only transmit epistemic properties from one subject S_1 to another S_2 ; so, if S_1 has no knowledge, then S_2 cannot acquire knowledge through S_1 's testimony. Simply put, it would be like relying on a friend to pay our bill, but that friend is broke as we are; if he can't pay the bill, we can't either. An analogy between testimony and memory is often presented as a reason for this premise (K). For instance, Dummett¹¹ writes that:

If remembering something is to count as retaining a knowledge of it, it must have been known when originally witnessed or experienced; if it was derived from a misperception or misapprehension, the memory cannot of course rank as knowledge. The same naturally applies to taking something to be so, having been told it: the original purveyor of the information – the first link in the chain of transmission – must himself have known it, and therefore have been in a position

¹¹ "Testimony and Memory," 264.

to know it, or it cannot be knowledge for any of those who derived it ultimately from him.

The main point has to do with the fact that memory only preserves knowledge from a moment to another, so testimony also only transmits knowledge from a sender to a receiver. In both cases no knowledge is generated, only preserved or transmitted. But is that plausible? A direct objection to (K) is presented by Peter J. Graham¹² through the following compelling counterexample:

TEACHER CASE: A devout creationist teaches at a public school where she must teach a section on evolutionary theory. She does not believe a word of it, but is a dedicated and responsible teacher. She develops a near expert understanding based on deep reading of books and articles on evolutionary science. She even develops a deep understanding of fossils that parallels highly skilled scientifically trained expertise. On a fieldtrip she discovers a fossil that proves that ancient humans [from which we evolved] once lived in this area (itself a surprising discovery no one knew before) [– for brevity let’s call this proposition *p*]. Though she does not believe it, when she tells this to her students, they believe her. Because of her commitment to teaching, her exposure to evolutionary science, and her mastery of fossils, she would not say what she did unless it were true. Her assertion is a reliable indicator. Relying on their teacher, the schoolchildren would not easily be mistaken.

In this case the sender, the creationist teacher, doesn’t know that *p*, because she doesn’t believe *p* (only accepts *p* for practical teaching and research purposes).¹³ However, the receivers, the students, know that *p*, given that they have a safe and justified true belief, un-Gettierized, that *p*. This case shows that a sender *S* doesn’t know that *p*, but a recipient *R* can come to know that *p* on the basis of the testimony that *p* provided by *S*. Moreover, it is shown that testimony not only transmits knowledge but can also *generate* it. This is because the first link in the chain of transmission, the teacher, doesn’t know that *p*; yet her testimony was able to generate knowledge in the students – since these students know

¹² “Testimonial Knowledge: A Unified Account,” *Philosophical Issues* 26, 1 (2016): 176, <https://doi.org/10.1111/phs.12082>.

¹³ Here we are claiming that *belief* and *acceptance* are different states. On the one hand, belief is an involuntary dispositional state, aims at truth, follows evidence, is ideally coherent, and comes in degrees. On the other hand, acceptance is voluntary, aims at pragmatic success, follows interests and desires, and allows for contradiction. See Jonathan Cohen, “Belief and Acceptance,” *Mind* XCVIII, 391 (1989): 367–89, <https://doi.org/10.1093/mind/xcviii.391.367>, Andrei Buckareff, “Acceptance and Deciding to Believe,” *Journal of Philosophical Research* 29 (2004): 173–90, https://doi.org/10.5840/jpr_2004_17, and Hamid Vahid, “Alston on Belief and Acceptance in Religious Faith,” *The Heythrop Journal* 50, 1 (January 2009): 23–30, <https://doi.org/10.1111/j.1468-2265.2009.00430.x>.

through the teacher's testimony something new that no one has ever known before.¹⁴ Thus, premise (K) is false.

Since (K) is false, REPORT CASE is not plausible to deny the view that group testimony is a reducible source of knowledge. For, just as students, in TEACHER CASE, can come to know a theory from the testimony of a teacher who does not know that theory (because she does not believe it), it is also possible that recipients of the public report, in REPORT CASE, can come to know that q on the basis of the testimony that q provided by a single member of the committee who does not know that q . In other words, since a sender S need not know that q in order for a recipient to acquire knowledge that q through S 's testimony, there is no reason to claim that lack of knowledge on the part of singular members of the committee, in REPORT CASE, precludes recipients from knowing that q through their testimony. On this basis we can state that REPORT CASE does not show that reductionism is false.¹⁵

3.2 Problems With the Second Argument

The second counterexample, TALK CASE, has advantages over REPORT CASE, since it is not based on the controversial premise (K). With regard to TALK CASE, Kallestrup¹⁶ argues that group testimony is not reducible to the testimony of any of its members, not even to the assistant member who is giving the talk. This is because if someone x testifies that q , then x properly asserts that q . And if x

¹⁴ This case has advantages over similar ones. One of the most popular cases is the creationist teacher presented by Jennifer Lackey, "Testimonial Knowledge and Transmission," *The Philosophical Quarterly* 49, 197 (1999): 471–90, <https://doi.org/10.1111/1467-9213.00154>; *Learning from Words: Testimony as a Source of Knowledge* (Oxford University Press, 2008). This teacher professionally teaches the theory of evolution, but does not believe in this theory and thus does not know it. But her students are in a position to gain knowledge about the theory of evolution. The problem is that Lackey case does not undermine (K), given that the first links in the chain of transmission, that goes back to Darwin, know that the theory of evolution is the case. This last teacher just skips a link in the chain of testimonial knowledge and so knowledge is not generated by the chain. However, in the case modified by Graham ("Testimonial Knowledge") we do not have this problem. For a discussion of such cases see Peter J. Graham, "Can Testimony Generate Knowledge?" *Philosophica* 78 (2006): 105–27, J. Adam Carter and Philip J. Nickel, "On Testimony and Transmission," *Episteme* 11, 2 (2014): 145–55, <https://doi.org/10.1017/epi.2014.4>, Stephen Wright, *Knowledge Transmission* (Routledge, 2018), <https://doi.org/10.4324/9781315111384>.

¹⁵ Lackey, "A Deflationary Account of Group Testimony." used a similar argumentation to rule out counterexamples with an identical structure to REPORT CASE. However, she did not deal with counterexamples with an identical structure to TALK CASE.

¹⁶ "Groups, Trust, and Testimony."

properly asserts that q , then x understands q . However, the assistant does not understand nor is able to grasp q ; in the words of Kallestrup, “he lacks the required expertise to grasp such a complex proposition.”¹⁷ Therefore, the assistant does not testify that q . Neither Dr. X nor Dr. Y can testify that q , given that they are not even aware of this conclusion q . Thus, the testifier is the group itself. Is this a good argument? There are several problems with this argument.

3.2.1 Testimony Can Generate Understanding

First of all, underlying this argument is the following premise:

- (U) If a sender S doesn't understand ϕ , then a recipient R cannot acquire understanding of ϕ on the basis of the spoken or written word about ϕ provided by S .

Understanding, along with knowledge, is an important type of cognitive achievement. Ideally we not only want to know things, but also to understand them. There are several types of understanding, but here we will focus mainly on the more common – the so-called *objectual understanding* – which takes the form of “ S understands ϕ ” where ϕ is a certain domain, subject matter, or a body of information.¹⁸ For example, “ S understands the theory of evolution.” It is typically accepted that a necessary condition for a subject S to understand something ϕ is to be able to offer an explanation of ϕ .¹⁹ In this regard, Jonathan L. Kvanvig²⁰ holds that “understanding requires the grasping of explanatory and other coherence-making relationships in a large and comprehensive body of information.” In the same line of reasoning, Wayne D. Riggs²¹ argues that understanding a domain “requires a deep appreciation, grasp, or awareness of how its parts fit together,

¹⁷ *Ibid.*

¹⁸ In addition to *objectual understanding*, one can also distinguish *interrogative understanding*, which take the form “ S understand *why* ϕ ”, and *propositional understanding*, which take the form “ S understand *that* p .” It can be argued that there is nothing distinctive about this latter kind of understanding, because propositional understanding can be reduced to propositional knowledge. See Emma C. Gordon, “Is There Propositional Understanding?” *Logos & Episteme* 3, 2 (2012): 181–92, <https://doi.org/10.5840/logos-episteme20123234>.

¹⁹ Since understanding somehow implies reflexively accessible bases in support of the object of understanding, Pritchard (*Epistemology* (Palgrave Macmillan UK, 2016), 128, <https://doi.org/10.1007/978-1-137-52692-2>) considers that understanding, unlike knowledge, “is of its nature an epistemically internalist notion.”

²⁰ *The Value of Knowledge and the Pursuit of Understanding* (Cambridge University Press, 2003), 192, <https://doi.org/10.1017/cbo9780511498909>.

²¹ “Understanding ‘Virtue’ and the Virtue of Understanding,” in *Intellectual Virtue* (Oxford University Press, 2003), 217, <https://doi.org/10.1093/acprof:oso/9780199252732.003.0010>.

what role each one plays in the context of the whole, and of the role it plays in the larger scheme of things.” Understanding also has a social dimension: in epistemic communities there are some members – such as epistemic authorities, experts, teachers – who help other members to better understand something through testimony. Thus, it seems that testimony plays a relevant role in spreading understanding within epistemic communities.

Given this, one might find it intuitive to accept premise (U). For, if the sender does not understand a particular topic or domain, how can the recipient understand that topic from the sender’s word? In other words, it does not seem possible to improve understanding of a specific domain in a recipient if the sender has no understanding of that domain. It is based on (U) that it can be said that, in relation to TALK CASE, if the assistant doesn’t understand the domain of things to which q belongs, then a recipient, an audience, cannot acquire understanding of that domain or body of information of which q is part through the assistant’s spoken or written word. However, given that the audience has acquired knowledge and understanding of the information set containing q , the assistant cannot be regarded as testifying q , but instead “the testifier is the group itself” as Kallestrup²² supports. Is this plausible? We want to argue that the intuition that underlies premise (U) is mistaken. In order to show this, to rule out premise (U), we have developed the following counterexample:

SCHOOL ASSISTANT CASE: Suppose the creationist teacher has already taught her students the theory of evolution. But because of an illness, she cannot teach the lesson in which she would explain a body of information that contains the topic that *a new fossil proves that ancient humans from which we evolved once lived in a certain area* – let’s call this information ψ for short. Since this teacher did not come to the class, because of her illness, a school assistant intentionally decided to give this lesson about ψ to these students. However, such a school assistant does not know nor understand anything about science, evolutionary theory or fossils. Nevertheless, she fetched the pedagogical notes about ψ from the teacher’s desk and began to read them rigorously so that these students could learn about ψ . Due to prior knowledge of the theory of evolution and fossils, such students were able to understand ψ .

In this case, the sender, the school assistant, does not understand ψ , but the recipients, the students, can acquire understanding of ψ from the written and spoken word provided by the assistant. Due to students’ background knowledge, it is reasonable to accept that they are able to explain ψ and make connections between ψ and the theory of evolution through the school assistant’s testimony. If

²² “Groups, Trust, and Testimony.”

so, it is shown that testimony can be a generative source of understanding. Just as this situation can occur in the individual testimony, can also occur in the group testimony. Thus, as it is possible for students to gain understanding of a phenomenon ψ from the spoken and written word provided by the school assistant who does not understand ψ , it is also possible, with regard to TALK CASE, for an audience of scientists to gain understanding of a domain of things to which q belongs from the spoken and written word provided by the research assistant who does not understand that domain of things. Given that, TALK CASE is not decisive in rejecting reductionism, because the assistant doesn't need to understand a domain (which contains q) in order to testify and generate understanding about that domain in his audience.²³

3.2.2 Proper Assertion and Understanding

Advancing another type of objection, it may be pointed out that Kallestrup²⁴ seems to accept in his argument that if the assistant does not understand q , then he cannot properly assert q . And if the act of testifying involves proper assertion, the assistant cannot testify that q .²⁵ However, one can reply by stating that, although the assistant does not understand q , he understands another proposition; namely, that the content q , whatever it may be, is the result of evidence provided by Dr. X and Dr. Y with the application of *modus ponens* (we use r to abbreviate this proposition). Thus the assistant can properly assert r and, based on this, an audience of scientists is able to understand and know both r and q through this assistant's testimony. In other words, the assistant believes r and this belief enables him to properly assert q , on the basis of which the audience may then understand q .²⁶ Yet this argument assumes that there can be no proper assertion without understanding. But is this true?

To make this clear, let's first look at a brief characterization of *assertion*. The speech act of assertion denotes the familiar phenomenon by which a subject states, reports, contends, or claims that something is the case. But what distinguishes assertion from other speech acts (such as speculations or guesses)? It is typically

²³ For a different argument against premise (U), see Federica Isabella Malfatti, "Can Testimony Generate Understanding?" *Social Epistemology* 33, 6 (2019): 477–90, <https://doi.org/10.1080/02691728.2019.1628319>.

²⁴ "Groups, Trust, and Testimony."

²⁵ This requirement that a *proper assertion* is necessary for *testimony* seems very strong. See Sanford C. Goldberg, "Assertion, Testimony, and the Epistemic Significance of Speech," *Logos & Episteme* 1, 1 (2010): 59–65, <https://doi.org/10.5840/logos-episteme20101121>.

²⁶ We are grateful to Amanda Bryant for this idea.

accepted, according to Sanford C. Goldberg,²⁷ that “assertion is the unique speech act that is governed by a particular rule: the so called norm of assertion.” Thus, the speech act of assertion can be individuated by reference to this rule or norm. Such norm has the following structure:

One should assert that p only if ϕ .

where we replace “ ϕ ” with the condition that captures the content of this norm. There has been a lot of disagreement over what is the most appropriate way to replace “ ϕ ”. The main candidates for ϕ are the following: one knows that p ,²⁸ it is true that p ,²⁹ one is epistemically certain that p ,³⁰ it is reasonable for one to believe that p .³¹ But the argument put forward by Kallestrup³² against reductionism assumes a different norm of assertion:

Understanding norm of assertion (UNA): One should assert that p only if one understands p .

It is based on this norm that Kallestrup claims that the assistant in TALK CASE is not properly asserting that q and, thus, is not testifying that q , since “nobody can properly assert a proposition they do not understand.”³³ But is this UNA rule plausible? We want to argue that UNA is false. On the one hand, if UNA is read in the sense of *objectual understanding*, as we are using in the previous section (which is the most typical sense of understanding),³⁴ then a speaker S should assert that p only if S understands a domain or body of information of which p is part. But this is a very strong requirement for assertion. A speaker need not understand a domain of things to which p belongs in order to make a proper

²⁷ *Assertion: On the Philosophical Significance of Assertoric Speech* (Oxford University Press, 2015), 3, <https://doi.org/10.1093/acprof:oso/9780198732488.001.0001>.

²⁸ Timothy Williamson, *Knowledge and Its Limits* (Oxford University Press, 2000), <https://doi.org/10.1093/019925656x.001.0001>.

²⁹ Matthew Weiner, “Must We Know What We Say?” *Philosophical Review* 114, 2 (2005): 227–51, <https://doi.org/10.1215/00318108-114-2-227>.

³⁰ Jason Stanley, “Knowledge and Certainty,” *Philosophical Issues* 18, 1 (2008): 35–57, <https://doi.org/10.1111/j.1533-6077.2008.00136.x>.

³¹ Lackey, *Learning from Words*.

³² “Groups, Trust, and Testimony.”

³³ *Ibid.*, 140.

³⁴ Here we will not address the reading of UNA with *propositional understanding*, because it is commonly argued – for example, by Gordon, “Is There Propositional Understanding?” – that propositional understanding is reduced to propositional knowledge. So with this reading UNA collapses into the knowledge norm of assertion (KNA). Cases such as TEACHER CASE can be presented as counterexamples to KNA. See Lackey, *Learning from Words*.

assertion that *p*. For instance, if our car breaks down, we can properly assert to our auto mechanic: “the car is not working properly.” And we can make this assertion in absence of understanding about the domain related to the functioning of cars and their mechanical components. It is not necessary to have any understanding about automobiles and their mechanical operation to make this assertion without being subject to criticism. If UNA were true, only experts in a given domain could make appropriate assertions about some proposition of that domain.³⁵ However, such requirement would absurdly preclude laypeople from making assertions. Our SCHOOL ASSISTANT CASE can also be a counterexample for UNA. For, in this case we have a school assistant who makes an assertion about fossils without any understanding about this domain. Moreover, her assertion does not seem improper, given that she is aware that the teacher’s notes she is using are appropriate, reliably conveying information for students.

On the other hand, if UNA is read in the sense of *interrogative understanding*, then a speaker *S* should assert that *p* only if *S* understands *why p*. And for *S* to understand *why p*, *S* must give an explanation or reason *q* why *p* (in other words, interrogative understanding is equivalent to understanding that *p* because *q*). According to Alison Hills,³⁶ “understanding why *p*, though, requires more than the correct belief that *p* because *q*. It requires a grasp of the reason why *p*, or more precisely, a grasp of the relationship between *p* and *q*.” This means that, for example, *S* is able to provide the correct explanation or reason *q* for the information that *p*, to draw the conclusion that *p* from the explanation that *q*, explaining the relation between *p* and *q* in his own words, among other aspects.³⁷ But requiring such an *interrogative understanding* to make proper assertions is a very strong demand. For example, we can adequately assert to our auto mechanic that the car we are using is not accelerating properly (compared to its usual operation), even when we are unable to find out a non-circular explanation or reason for it. And if our auto mechanic tells us that the car is not accelerating properly because the car has bad spark plugs, we can thus gain knowledge of such an explanation, yet we may be in a situation where we do not understand such an explanation. This is because, unlike the auto mechanic, we don’t have any background beliefs about spark plugs nor are we able to grasp how the spark plugs are associated with the operation of car acceleration. However, there seems to be

³⁵ For the idea that what distinguishes experts from laypeople is at least their understanding of a domain, see Michel Croce, “On What It Takes to Be an Expert,” *The Philosophical Quarterly* 69, 274 (2018): 1–21, <https://doi.org/10.1093/pq/pqy044>.

³⁶ “Understanding Why,” *Nous* 50, 4 (2015): 663, <https://doi.org/10.1111/nous.12092>.

³⁷ For a complete characterization see Hills, “Understanding Why.”

nothing wrong with making that initial assertion about car acceleration problems. Once again, to require understanding to make proper assertions would be to improperly intellectualize the speech act of making assertions, since only experts would be entitled to make assertions. Even those who are in favor of accepting the condition of interrogative understanding in some cases consider that “the proposal that understanding is the norm of assertion is false,” such as J. Adam Carter and Emma C. Gordon³⁸ in claiming that:³⁹

There are indeed many proper assertions for which any kind of understanding is not a necessary condition – for example, ‘the bird is yellow’ (an assertion of perceptual knowledge) does not require being able to fill out any further claim including the word ‘because’ before that particular assertion counts as permissible.

What is the main epistemic aim of assertion? As we argued above, this main purpose does not seem to be related to the speaker’s understanding; instead it seems more plausible to accept that the relevant aim of assertion is generating (or at least it has the disposition to generate) some epistemic status in the hearer or recipient. This is for two main reasons: First, because the social function of language is to convey or communicate information, often through assertion speech acts. Second, because we are cognitively limited beings (that is, we cannot afford to know many things firsthand or in isolation), we need to rely on other people’s words, especially on proper assertions, to gain knowledge. Based on similar reasons, Charlie Pelling⁴⁰ argues for the following rule of assertion:⁴¹

Audience-oriented norm of assertion (ANA): One should assert that *p* only if it is fit to give a hearer knowledge that *p*.⁴²

³⁸ “Norms of Assertion: The Quantity and Quality of Epistemic Support,” *Philosophia* 39, 4 (2011): 631, <https://doi.org/10.1007/s11406-011-9317-6>.

³⁹ Mona Simion, “The Explanation Proffering Norm of Moral Assertion,” *Ethical Theory and Moral Practice* 21, 3 (2018): 486, <https://doi.org/10.1007/s10677-018-9922-6> also rules out UNA.

⁴⁰ “Assertion and the Provision of Knowledge,” *The Philosophical Quarterly* 63, 251 (2013): 294, <https://doi.org/10.1111/1467-9213.12013>.

⁴¹ Manuel García-Carpintero, “Assertion and the Semantics of Force-Markers,” in *The Semantics/Pragmatics Distinction*, ed. Claudia Bianchi (CSLI Publications, 2004), 156 developed an assertion rule very similar to this one.

⁴² It is necessary to clarify how this rule allows to overcome those counterexamples of proper assertions without hearers, such as a private assertion (for example, written in the secret diary). According to Pelling (“Assertion and the Provision of Knowledge,” 300.), “we should think of assertions which are fit to give knowledge as those that, even if they do not in fact give knowledge, at least have the right kind of evidential bases – the kind which they would need to have in order to give knowledge.”

In a similar way, Christoph Kelp⁴³ holds that assertion has the epistemic function of generating knowledge in hearers, defending the following rule of assertion:

Functionalist norm of assertion (FNA): One should assert that *p* only if it has the disposition to generate knowledge that *p* in hearers.

If these ANA and FNA assertion rules are plausible, we can say that in TALK CASE the assistant made a proper assertion. This is because what the assistant utters about *q* to his audience in that context has the disposition to generate knowledge and understanding in his audience about *q*, even if the assistant himself has no knowledge or understanding of that domain. After all, what the assistant utters comes from reliable processes (applying *modus ponens* from information received from Dr. X and Dr. Y) and this utterance about *q* generates knowledge and understanding in his audience of scientists. He does not violate any plausible and relevant assertion rule that we are considering.⁴⁴ But, if this is so, then there is no reason to hold that the assistant is not really testifying that *q*. Thus, TALK CASE does not show that reductionism in relation to group testimony is false.

3.2.3 Assertion and Grasping

Perhaps Kallestrup⁴⁵ can reply that the biggest problem is not with the violation of an assertion norm, but with the act of asserting itself. In other words, the assistant's speech act cannot even be considered an assertion because he does not *grasp* the proposition under consideration; in Kallestrup's own words, the assistant "even lacks the required expertise to grasp such a complex proposition."⁴⁶ So it seems that Kallestrup⁴⁷ is committed to the idea that *grasping* is a necessary condition for assertion and, since the assistant is not able to *grasp* the proposition *q*, he cannot believe and assert such a proposition or testify it.

However, Kallestrup reasons are not plausible; for, as we will argue below, on the one hand, *grasping*, understood in a phenomenal sense, is not a necessary condition for believing and asserting; and, on the other hand, there are other

⁴³ "Assertion: A Function First Account," *Noûs* 52, 2 (2016): 16, <https://doi.org/10.1111/nous.12153>.

⁴⁴ It could be argued that what the assistant utters does not satisfy the knowledge norm of assertion (KNA). However, with TEACHER CASE it can be indorsed that KNA is false. See Lackey, *Learning from Words*, 111–14 and Pelling, "Assertion and the Provision of Knowledge," 305–7.

⁴⁵ "Groups, Trust, and Testimony."

⁴⁶ *Ibid.*, 140.

⁴⁷ "Groups, Trust, and Testimony."

theories of *grasping* in which we can hold that the assistant *grasps* such a proposition *q*. So it is worth questioning: what *grasping* a proposition is? Given that Kallestrup does not give us any account of “grasping,” it is important to survey what are the main theories of grasping.

Before this analyses, in order to show that Kallestrup is wrong, we will start by presenting simple and intuitive examples in which a subject can genuinely assert that *p*, for example by believing and reporting that *p*, without *grasping* (in a phenomenal sense) that *p*. If these examples are plausible, we show that there can be assertion without grasping (in such a phenomenal sense). Let’s look at two cases:

COLOR-BLIND CASE: Suppose that Mary is a complete color-blind or achromat; so, she is unable to grasp the proposition that <ripe tomatoes are red>. However, she learned this proposition at school, by trusting her teacher testimony, and believes that it is so. For this reason, she is able to assert that <ripe tomatoes are red>, when reporting this to her son, without really grasping this proposition.

PRE-TEEN CASE: Suppose that Joseph, a pre-teen, has not yet acquired the required expertise to grasp a complex scientific proposition, such as <the Sun is about 1,300,000 times the volume of the Earth>. But, he learned this proposition from his father, an astrophysicist, and by trusting his father he believes that proposition is true (although he doesn’t grasp it). When he is talking to his friends, he is able to assert that <the Sun is about 1,300,000 times the volume of the Earth>, by reporting it, without grasping this proposition.

These examples illustrate that there is a state of mind – *grasping* (related with phenomenal experiences) – that goes beyond belief, knowledge and assertion, in such a way that one can assert that *p* (as well as one can believe and know that *p*) without grasping that *p*. Thus, if such cases illustrate that there can be assertion without grasping, the same can happen with the assistant in TALK CASE. In fact, the assistant knows that proposition *q* is not meaningless; moreover, he knows that *q* belongs to the scientific domain, and he also knows that *q* was produced by a reliable process through the application of an inference of *modus ponens* from the information received by the reliable testimonies of Dr. X and Dr. Y. If it is so, the assistant can assert that *q*, when reporting such proposition to his audience, even if he is unable to grasp such a complex proposition (similarly to what happens in COLOR-BLIND CASE and in PRE-TEEN CASE).

In such cases, it seems that the subjects under consideration genuinely make an assertion. Following Goldberg,⁴⁸ we can say that they are uttering a declarative sentence, in a sincere way (and not pretending or guessing), aiming communicate knowledge (which they received by trusting on reliable testimonies). In addition,

⁴⁸ *Assertion*, 6–9.

the assertions of such subjects can be challenged. For example, a recipient can ask on what basis they make these claims – and they can respond that they make these claims on the basis of reliable testimonies (in conjunction with an inference rule, at least in TALK CASE).⁴⁹ This can be considered as evidence that these subjects can have for their claims.

But although they assert a proposition, there is a sense in which we can say that they do not grasp that proposition. Why not? According to David Bourget,⁵⁰ the best theory of grasping is the *phenomenal theory*, in which “grasping a proposition is a matter of having a phenomenal experience that has the proposition as its content,” namely:

(TFT) The Phenomenal Theory: To grasp *p* is to have a phenomenal experience with *p* as content.

In such theory, experiences have intentional contents in the weak sense of presenting things. For example, when I look at one of my books, I seem to undergo an experience that *presents* me with a rectangular shape with many sheets of paper, etc. (TFT)’s central point can be summed up in the idea that *grasping is a matter of experience*, in such a way that if we cannot experience the content of a proposition, then we don’t grasp it. That’s why Mary, in COLOR-BLIND CASE, doesn’t grasp the proposition that <ripe tomatoes are red>; precisely because she lacks the ability to experience the color red (in perception or imagination). But suppose it is possible to perform a surgical operation that allows Mary to gain the ability to see colors for the first time. In that case, Mary would be able to grasp the proposition under consideration.

Something similar can be said about PRE-TEEN CASE. In this case, Joseph doesn’t grasp the proposition that <the Sun is about 1,300,000 times the volume of the Earth> because he is unable to experience this content or something similar; for example, because he has no visual model available to help him in such task. The same is true with TALK CASE; the assistant doesn’t grasp the proposition that *q*, given that he lacks the required expertise to build a visual model (or do not have a model available) that would allow him to experience such content. Despite this, even though they do not have this type of grasping, as we argued above, they can assert the proposition in question. However, suppose that Joseph’s father, in PRE-TEEN CASE, builds a model to assist his son in grasping the new proposition

⁴⁹ But suppose the recipient is not satisfied with that answer. According to Goldberg (*ibid.*, 75), in this situation these subjects are epistemically entitled to pass the epistemic burden to those people from whom they received those beliefs.

⁵⁰ “The Role of Consciousness in Grasping and Understanding,” *Philosophy and Phenomenological Research* 95, 2 (2015): 19, <https://doi.org/10.1111/phpr.12208>.

learned. Such a model draws, for example, an analogy that allows him to visualize that the relationship between the volume of the sun and the earth is like the relationship between the volume of a basketball and the volume of an apple seed.⁵¹ This model helps Joseph to grasp this content by providing him with a visual experience. A similar model could help the assistant, in TALK CASE, to grasp the complex proposition *q*. But the assertion can occur regardless of this type of grasping.

There are other theories of grasping in which it is not even necessary to offer a visual model or some kind of experience for the subject to grasp the proposition in question. The most typical and influential theory of grasping is the *inferential theory*. On this theory, following Bourget, “grasping a proposition is a matter of having a thought that represents the proposition and is suitably connected to other mental states through inference-like dispositions.”⁵² So, grasping a proposition *p* has to do with being able to reason properly about the implications between *p* and other propositions. Such a theory is supported by Kvanvig,⁵³ Martine Nida-Rümelin,⁵⁴ Stephen R Grimm,⁵⁵ Daniel A. Wilkenfeld⁵⁶ and can be presented like this:

(TIT) The Inferential Theory: A thought *t* with content *p* is a grasping of *p* to the extent that *t* is appropriately inferentially connected to other mental states of the subject.⁵⁷

But this theory allows us to conclude that, in the cases mentioned above, as well as in the TALK CASE, the subjects really grasp the propositions under consideration, precisely because they are able to make inferences based on such propositions. Starting with COLOR-BLIND CASE, even if Mary is unable to have *phenomenal experiences of red* (call it *R*), she can learn (from reliable testimonies) and know that <red things cause *R* experiences>; and based on that she can infer that <ripe tomatoes cause *R* experiences>, <at typical vertical traffic lights, the lamp at the top, when switched on, causes *R* experiences>, <when a fire truck

⁵¹ This example is inspired by Bourget (*ibid.*, 3–4).

⁵² *Ibid.*, 11.

⁵³ *The Value of Knowledge and the Pursuit of Understanding*.

⁵⁴ “Grasping Phenomenal Properties,” in *Phenomenal Concepts and Phenomenal Knowledge* (Oxford University Press, 2007), 307–38, <https://doi.org/10.1093/acprof:oso/9780195171655.003.0013>.

⁵⁵ “Understanding,” in *The Routledge Companion to Epistemology* (Routledge, 2011), 110–20.

⁵⁶ “Understanding as Representation Manipulability,” *Synthese* 190, 6 (2013): 997–1016, <https://doi.org/10.1007/s11229-011-0055-x>.

⁵⁷ See Bourget, “The Role of Consciousness in Grasping and Understanding,” 14.

appears, it causes *R* experiences>, <things marked with the RED symbol (in the ColorAdd system) cause *R* experiences>, and so on.

Something similar can be said about PRE-TEEN CASE. Joseph can establish numerous inferences from the proposition <the Sun is about 1,300,000 times the volume of the Earth>, such as <the sun is larger than our school>, <the sun cannot be a soccer ball>, and so on. This same strategy is available to the assistant in TALK CASE. In such a case from the complex proposition *q*, and given his background knowledge, the assistant can infer that <*q* belongs to the scientific domain>, <*q* is not about philosophy or literature>, <*q* must be presented to an audience of scientists and not to an audience of children>, and so on. So, if the assistant is able to make inferences based on the proposition under consideration, then according to (TIT) theory he really grasp the proposition *q*.

It is worth considering a last theory of grasping that is based on the work of Timothy Williamson.^{58,59} On Williamson's account, to grasp a word or proposition is to be a member of a community that uses that word or proposition. In addition, one counts as a member of a community insofar as one participates in relevant causal interactions with other members of that community. In this regard, Williamson⁶⁰ holds that such members "use a word as a word of a public language, allowing its reference in their mouths to be fixed by its use over the whole community." In turn, a subject does not grasp a word or proposition when, following Williamson, there is a "lack of causal interaction with the social practice of using that word" or proposition.⁶¹ Simply put, this theory holds that:

(TET) The Externalist Theory: To grasp *p* is to be a member of a community that uses *p*.

Based on this theory we can claim that the subjects in the cases above grasp the propositions under consideration, given that they have a "sufficiently fluent engagement in the practice" in which such propositions are used. For example, Mary is a member of a community through which she has causal interactions (for example, through testimony) to use the proposition that <ripe tomatoes are red>. This is also true of Joseph regarding the proposition that <the Sun is about 1,300,000 times the volume of the Earth>. And the same solution can be applied to the assistant in TALK CASE. In such a case it can be argued that the assistant has

⁵⁸ "Conceptual Truth'," *Aristotelian Society Supplementary Volume* 80, 1 (2006): 1–41, <https://doi.org/10.1111/j.1467-8349.2006.00136.x>; *The Philosophy of Philosophy* (Blackwell Publishing Ltd, 2007), <https://doi.org/10.1002/9780470696675>.

⁵⁹ We are grateful to Bruno Jacinto for this suggestion.

⁶⁰ "Conceptual Truth'," 36.

⁶¹ *Ibid.*, 38.

causal interaction with the social practice of using the proposition q (after all, he acquired such proposition q for being part of a research community and for giving a talk to an audience of scientists who are going to use this proposition). Thus, contrary to Kallestrup,⁶² we can conclude that there is a sense in which the assistant grasps the proposition q . But even in an interpretation in which he does not grasp q , as seen in the phenomenal theory, the assistant can continue to assert q .

3.2.4 A Final Objection

As a last objection, it can be argued that the assistant in TALK CASE, because he does not understand complex propositions about physics such as p and complex mathematical propositions such as $p \rightarrow q$, is not able to apply a *modus ponens* based on those propositions.⁶³ For it does not seem that one can properly apply *modus ponens* from what is not understood. For example, suppose we received from two different people two identical sentences containing indexicals; if we do not know the context of illocution of such sentences, it does not seem that we can say that they express the same proposition, nor can we properly use such sentences to make inferences. Thus, if the assistant does not understand the content of complex propositions, how can he know that the proposition provided by the physicist is the same proposition that corresponds to the antecedent of a conditional provided by the mathematician? The point is that the assistant could properly apply *modus ponens* rule only if it were just a syntactic application. But it does not seem that applying the *modus ponens* rule is just a syntactic application. Instead, in order to properly apply *modus ponens* one must know the semantics, the meaning, of what is being considered.

A proponent of the non-reductionist view may be able to resist this latter objection by slightly modifying TALK CASE. Suppose, instead of the assistant applying *modus ponens*, he has a different and much simpler task. Dr. X and Dr. Y gave the assistant a three-part pre-written text: the first part only has the conclusion p and a blank space to put the proof provided by Dr. X; the second part only has the conclusion $p \rightarrow q$ and a blank space to put the proof provided by Dr. Y; in the last part appears the application of *modus ponens* with the final conclusion q . Suppose further that Dr. X and Dr. Y have no other communication with each other and are able to reach desired individual conclusions without the other knowing this result. Thus, the assistant only needs to perform the following

⁶² "Groups, Trust, and Testimony."

⁶³ We are grateful to Ricardo Santos for this idea.

task: if the assistant independently receives Dr. X proof and Dr. Y proof, he will mechanically fill in the blanks of this pre-written text with those proofs and read the entire text to his audience.⁶⁴

However, this modification of TALK CASE is still worse than the case originally presented in section 2. For in this modified case the assistant has no cognitive achievement. At least in the original case the assistant had a relevant cognitive task, making an important contribution: trying to apply *modus ponens* (which seems to involve not only syntactic but also semantic tasks). Now his role is reduced to a mere automatic input and output system, working only at syntactic level. There is no salient difference between this assistant or a computer in performing the task in question. If this is so, then it is not clear that we are facing a group testimony case, because testimony involves intentions to convey information and it is doubtful that an automatic system, like a computer, has such intentions. Furthermore, if what is transmitted by the system results from a simple automatic merging of individually written and non-collaborative information from Dr. X and Dr. Y, then this does not seem to be a group testimony case, but rather a case of a single output containing several instances of individual testimony. But if the group is effectively collaborative and moreover if the assistant is not like a computer and really intends to convey information on group behalf to a particular audience then, as we argued in the previous sections, the testimony of such a group can be reducible to the assistant's testimony.

4. Advantages of Reductionism

In the previous sections we argued that premises (K) and (U) are false. A more plausible and reasonable epistemic principle to the phenomenon of testimony seems to be the following:

- (R) If a sender *S* is not reliable in transmitting information that *p*, then a recipient *R* is not in position to know (or to understand) that *p* on the basis of the testimony that *p* provided by *S*.

This principle (R) allows testimony to function as a generative epistemic source. This is because, as we saw in TEACHER CASE, a sender knowing that *p* is not necessary for a recipient to acquire testimonial knowledge that *p*; rather, what is necessary, among other conditions, is reliability in the transmission of information by the sender. Without reliable transmission, there is no testimonial knowledge acquisition. A similar point can be made regarding understanding.

⁶⁴ This modification of TALK CASE is closer to the original case presented by Bird, "Social Knowing," 34–35.

Looking at REPORT CASE again, there is no problem that no single member of committee knows that q , because a sender's knowledge that q is not necessary for a recipient to acquire knowledge that q on the basis of sender's testimony. The important thing, according to (R), is the reliability in the transmission of information. In this regard, a singular member S of committee, being a competent scientist, can individually and reliably transmit the information that q . Thus, the source of testimonial knowledge is not the committee itself, but it is somehow reducible to the reliability of S . For, whether recipients acquire knowledge that q on the basis of S 's testimony depends on whether S is reliable in transmitting information that q . An analogous description can be made for TALK CASE.

This is a reductive view on group testimony, since the epistemic status of testimony of a group g on a given proposition p *reduces* to the reliability in the transmission of information that p by a singular member of g . Following this reductionist view, group testimony can be treated just like an instance of individual testimony, there being no special epistemology to deal with the phenomenon of group testimony. Thus, in order to explain and evaluate group testimony, we only need to use the available resources we have to evaluate individual testimony. From this perspective, the phenomenon of group testimony does not involve any mystery and can be explained in a parsimonious way, because it does not require new theoretical resources. However, following a non-reductionist view, there are several difficult explanatory tasks: in particular, it would be necessary to develop a successful framework to explain how groups can act and have intentions in a way that does not depend solely on the individual acts and intentions of its members. The reductionist view is simpler and is not subordinate to this explanatory framework.⁶⁵

⁶⁵ Acknowledgements: I am grateful to Ricardo Santos, Michel Croce, Amanda Bryant, Bruno Jacinto, Bogdan Dicher, David Yates, Célia Teixeira, Vitor Guerreiro, Sofia Miguens, Charles Travis, and Jennifer Lackey for helpful comments and discussion on an earlier version of this paper. I also thank audiences for their questions on the presentations of this work at: *The Epistemology of Testimony: Authority and Disagreement* (held at the University of Porto) and *LanCog Workshop on Group Epistemology* (held at the University of Lisbon). Any errors or omissions are my responsibility. Work for this paper was supported by the post-doctoral project CEECIND/01066/2017 of the Portuguese Foundation for Science and Technology.

MORAL REALISM AND THE PROBLEM OF MORAL ALIENS

Thomas GRUNDMANN

ABSTRACT: In this paper, I discuss a new problem for moral realism, the problem of moral aliens. In the first section, I introduce this problem. Moral aliens are people who radically disagree with us concerning moral matters. Moral aliens are neither obviously incoherent nor do they seem to lack rational support from their own perspective. On the one hand, moral realists claim that we should stick to our guns when we encounter moral aliens. On the other hand, moral realists, in contrast to anti-realists, seem to be committed to an epistemic symmetry between us and our moral aliens that forces us into rational suspension of our moral beliefs. Unless one disputes the very possibility of moral aliens, this poses a severe challenge to the moral realist. In the second section, I will address this problem. It will turn out that, on closer scrutiny, we cannot make any sense of the idea that moral aliens should be taken as our epistemic peers. Consequently, there is no way to argue that encountering moral aliens gives us any reason to revise our moral beliefs. If my argument is correct, the possibility of encountering moral aliens poses no real threat to moral realism.

KEYWORDS: disagreement, conciliatory view, steadfast view, moral realism, relativism

Robust moral realists believe that moral judgments have truth-values (cognitivism), that many of them are in fact true (denial of error theories) and that they are made true by facts that do not depend on moral attitudes of the subject or on moral codes of the subject's community (metaphysical realism). According to them, moral facts are completely objective and mind-independent. Moral realists can accept that there is moral disagreement of the everyday kind. Some of it is rooted in disagreement about non-moral facts and can be resolved by supplying further information about the case at hand. For example, two people may disagree about whether it is morally permissible for an agent to take home some item simply because they disagree about whether the agent owns this item. Another example: Descartes would disagree with many of us today about whether it is permissible to treat animals like things because he, in contrast to us, believes that animals do not have a mind and cannot suffer. Apart from this kind of disagreement, there is also genuine moral disagreement about moral values and duties or about how they apply to particular cases. Some such disagreements can be resolved by exchanging reasons. In other cases, the disagreement survives persistent exchange of reasons. Whether abortion is morally permissible, whether the air force is morally

permitted to shoot down an aircraft that has apart from its many innocent passengers a number of terrorists on board who are planning to steer the plane into a building with a huge number of residents, or whether we are morally permitted to perform active euthanasia in specific cases, seems to remain highly controversial even after all relevant arguments have been exchanged. In these rather local cases, moral realists may be tempted to suspend judgment.

As I said, moral realists can accommodate all these cases. But what happens when they encounter people who disagree with them in a more radical way? What should our rational reaction be when we meet people who persistently disagree with us not only about local moral issues but more broadly about the moral assessment of nearly every particular case and, more fundamentally, about moral principles as well? If these people appear to us as being intelligent, thoughtful, well-informed and if their views are without any obvious mark of incoherence, I will refer to them as “moral aliens.” One is inclined to say that one should not simply give in to those moral aliens but stick to one’s guns. However, how can the moral realist accommodate this intuition? From the moral realist’s perspective, only one of the disagreeing parties can be correct. If the moral alien can reasonably be taken to share all of our information about the non-moral facts and appears as internally rational and coherent as we are, then it seems that we must attribute similar weight to both perspectives. The resulting massive belief revision or even moral skepticism is in tension with our intuition that we should rationally stick to our guns when we encounter moral aliens. This seems to raise a severe problem for moral realism.

In this paper, I will argue that moral realists can solve this problem. More specifically, I will explain why moral aliens do not put any rational pressure on moral realists to revise their meta-ethical beliefs. Although this solution is motivated by deeply epistemological considerations which also apply to non-moral domains, it will turn out that it neither depends on a specific meta-epistemological position nor on a specific account of the epistemic significance of disagreement. In section I, I will explain and motivate the problem of moral aliens (for moral realism). In section II, I will argue that this problem dissolves when we look more carefully into its underlying epistemology. I will conclude with some general remarks about my solution to the problem.

1. The Problem of Moral Aliens for Moral Realists: The Deeper Motivation

The above sketched problem for moral realism can be articulated by the following inconsistent quartet of *prima facie* plausible assumptions:¹

- I. Moral aliens are possible.
- II. If one were to encounter (a significant number of) moral aliens, one would not be rationally required to revise one's moral judgments.
- III. If moral realism is correct, one would be rationally required to revise one's moral judgments significantly, if one were to encounter (a significant number of) moral aliens.
- IV. Moral realism is correct.

To put it in a nutshell: there is a genuine possibility with respect to which our intuitive judgment runs counter to what moral realism requires. This raises a severe problem for moral realism.

In what follows, I will explain and motivate each of the above assumptions. Let me start with (1). What are moral aliens? Being a moral alien is a relative property. Person A is a moral alien with respect to person B if A's moral perspective conflicts radically with B's moral perspective. Being such an alien does not imply anything about who is right and who is wrong.² It simply involves global (or at least widespread) moral disagreement at the level of particular moral beliefs and fundamental moral disagreement at the level of moral principles.³ However, this radical moral disagreement is not sufficient to constitute moral alienness. The

¹ Kieran Setiya, *Knowing Right From Wrong* (Oxford: Oxford University Press 2012) and Katia Vavova, "Moral Disagreement and Moral Skepticism," *Philosophical Perspectives* 28 (2014): 302-333, have also addressed the ethical significance of moral aliens. But they take a different stance on it. Setiya also starts with the intuition that we should not give in to what I call 'moral aliens'. But he derives very different metaethical consequence from my. He argues for some kind of epistemic externalism about moral evidence and disputes conciliatory views as well as a methodology of reflective equilibrium for ethics. Vavova argues that conciliationists would not be forced into moral skepticism if they were to encounter moral aliens since they could not reasonably assess them as epistemic peers. Although my own argument is close to hers, it differs in two significant respects. First, Vavova does not take the skeptical worry as a challenge to robust moral realism. Second, I fend off the skeptical challenge from moral aliens much more broadly than she does.

² The creature which I call "moral alien" is designed along the lines of what Setiya (*Knowing Right From Wrong*, 19-20) calls a "moral monster." However, while Setiya's terminology already involves an assessment of who is right and who is wrong, mine is neutral.

³ If only either the particular judgments or the accepted moral principles differ radically, the opponent will have less coherent beliefs than we have.

disagreement also has to persist under full disclosure of reasons and arguments to both parties. Moreover, the moral alien has to share all her non-moral beliefs (and evidence) with the person she is alien to. Finally, the alien must have a seemingly coherent system of moral beliefs that at least appears rational to her opponent. Here is then a definition of being a moral alien:

A is a moral alien with respect to B if and only if

- (i) A's moral beliefs conflict widely and fundamentally with B's moral beliefs,
- (ii) the disagreement between A and B persists through full disclosure of relevant evidence and relevant beliefs,
- (iii) A and B share their evidence and non-moral beliefs,
- (iv) A's moral position is without any obvious incoherence and appears rational to herself.

What kind of moral perspective would a moral alien have? Of course, a moral perspective could radically diverge from ours in many different respects. Let us start with a minimal characterization of the common core of "our" moral perspective: We take it that there is at least a *prima facie* moral duty (or a moral value) of minimizing suffering, of treating all humans equally, of not killing other people, of saving the survival of humanity etc. We also believe that justice matters morally. These seem to be some platitudes of our common moral perspective. Moral aliens would not only disagree with us about them but also consistently apply their shocking platitudes to particular cases. Moral aliens would, e.g., claim "Suffering is morally irrelevant" or "The white race has a supremacy over all other human races" or "Justice is morally irrelevant" or "We need not care about the survival of humanity" etc. Even more radical deviations from our moral point of view are intelligible. We might, e.g., encounter a moral alien who claims:

(B1) It is morally required to maximize human suffering.

Of course, one might be worried that this alien does not talk about what is morally required *in our sense of the word* when her moral platitudes are deviant to such an extent from ours. According to one popular view of moral concepts,⁴ they are determined by the moral platitudes which the subject associates with them. If two subjects associate two radically different sets of platitudes with the same linguistic or mental vehicle, it cannot instantiate the same concept. Hence, the creature we take to be a moral alien is not really disagreeing with us about moral

⁴ See Michael Smith, *The Moral Problem* (Oxford: Blackwell Publishing, 1996); Frank Jackson, *From Metaphysics to Ethics: A Defense of Conceptual Analysis* (Oxford: Oxford University Press, 1998).

issues, but disagrees with us merely verbally. Here is a passage from Jackson (*From Metaphysics to Ethics*, 132) along these lines:

Genuine moral disagreement, as opposed to mere talking past one another, requires a background of shared moral opinion to fix a common, or near enough common, set of meanings for our moral terms. We can think of the rather general principles that we share as commonplaces or platitudes or constitutive principles that make up the core we need to share in order to count as speaking a common moral language.

In this paper, I don't want to engage with the intricate debate about the nature of moral concepts, i.e., with the question whether these concepts should be understood in an internalist, externalist, or primitivist way.⁵ For the aim of the paper, this is not required. One can make moral aliens intelligible by telling a story about what motivates their diverging perspective. Recall (B1) which claims that it is morally required to maximize human suffering. Suppose that Theophilus believes that an act is morally required if and only if God commands us to do this thing.⁶ Theophilus also believes that God has told him to do his best to increase human suffering in order to punish humans for their constant sinning. Given this background and the rationality of Theophilus' non-moral beliefs, it seems even rational for Theophilus to believe (B1). One might object that this is not a persuasive example of a moral alien, because Theophilus has beliefs about non-moral facts (e.g., about the revelation of God's will) that conflict with ours and our related evidence. Here is another case that may better fit this requirement: Physiophilos has the strong moral intuition that the preservation of biological diversity on earth is an absolute moral obligation. Physiophilos knows that human actions have caused a massive reduction of this diversity in the past and also has good reason to believe that humans will continue with this if their population is not massively reduced. He believes that killing people is the most effective means to stop this practice. Physiophilos therefore concludes (B2) that there is the moral duty to kill the majority of humans. These cases may illustrate that moral aliens holding to moral platitudes that radically differ from ours are indeed possible.⁷

⁵ Whereas semantic internalists claim that the reference of moral concepts is determined by the platitudes related to them (e.g., Smith, *The Moral Problem*), semantic externalists claim that the reference of moral concepts is determined by their causal regulation (e.g., David Brink, *Moral Realism and the Foundations of Ethics*, Cambridge, MA: Cambridge University Press 1989). Primitivists (e.g., G.E. Moore, *Principia Ethica*, Cambridge, UK: Cambridge University Press 1903) claim that moral concepts such as "good" are not analyzable.

⁶ See G.E.M. Anscombe, "Modern Moral Philosophy," *Philosophy* 33 (1958): 1-19, for the modern debate.

⁷ Even if dispositions to believe determine meanings, these dispositions can be blocked to

The problem for moral realism does not depend on the actuality of moral aliens. Surely, we do not encounter them in everyday life. People whose behavior radically deviates from what we take to be morally permissible or required are typically either immoral or driven by their emotions and affects. But there seem to exist rare examples of people who at least approximate moral aliens. Adolf Hitler was an at least partly intellectually motivated racist. However, he relied on views about human races that are empirically false. In his infamous 1943 Posen speech, Heinrich Himmler encouraged SS-officers to fulfill their moral duty of annihilating the Jewish people in a morally decent way. His speech seems to express a moral view that radically differs from ours.

What about the second assumption of the inconsistent quartet?

- (2) If one were to encounter (a significant number of) moral aliens, one would *not* be rationally required to revise one's moral judgments.

(2) expresses a strong intuition about what is rational when we encounter people with radically different moral views. Kieran Setiya (*Knowing Right From Wrong*, 19–20, my italics) articulates this intuition in the following way:

For the first time, you meet a stranger. He agrees with you outside of ethics, but when it comes to practical reason, his beliefs are shocking. Fill in the details as you like. Perhaps he thinks we should be utterly selfish, that we should maximize aggregate happiness, no matter who is trampled on the way. It turns out that he, too, belongs to a homogeneous community, exactly as numerous as your own. What should you now believe? (...) *We should not defer to moral monsters but condemn them, however numerous they are.*

There are some uncontroversial truisms about situations like the one described: When we encounter someone who believes that we should maximize human suffering or that it is morally required to kill the majority of the human population, this is not only shocking news but also disgusting and repulsive for us. Moreover, our conflicting moral views might be so deeply entrenched in our perspective and so tightly related to our personality that our immediate reaction to the moral alien (whom Setiya—pejoratively—calls “moral monster”) is outright rejection. This much should be uncontroversial. What Setiya claims about these cases is much stronger. According to him, we do not only react in an adverse manner but are also *rationally required* to behave in this way. Setiya maintains

manifest in linguistic behavior under stimulus conditions by other beliefs. I take it that Williamson-like cases (Timothy Williamson, *The Philosophy of Philosophy*, Oxford: Blackwell Publishing, 2007) can establish at least this much. For example, Williamson argues that one might even reject the Proposition *All vixens are vixens*, if one believes that there are no vixens and that sentences with universal quantifier phrases have existential implications.

that “we *should* not defer to moral monsters but condemn them” (my italics). I agree that we have the intuition that it would be irrational for us either to give in to the moral alien or to suspend judgment upon encountering her.⁸

Let us move on to the third proposition of the quartet:

- (3) If moral realism is correct, one would be rationally required to revise one’s moral judgments significantly, if one were to encounter (a significant number of) moral aliens.

What is the motivation behind this assumption? From the moral realist’s perspective, one of us—either we or the moral alien—has formed a false moral belief. However, when we reflect on the epistemic situation, each party has something similar going for it: for us, our moral point of view looks rational; the same is true for them, from their perspective. Given this epistemic symmetry, belief revision towards the middle seems to be the only rationally permissible reaction for both parties. And this reaction is in tension with what (2) claims.

Let me present this motivation in more detail by presenting two complementary arguments that call for the realist’s revision of moral judgment. When we are confronting (a group of) moral aliens, they either share all our morally relevant evidence or they possess evidence different from ours on their side. First, consider the case in which our shared non-moral evidence comprises the morally relevant evidence. We can make sense of such a case by assuming that the only evidence for our moral beliefs is evidence of mundane descriptive facts. For example, we justify our belief that a particular action is morally wrong by referring to nothing but the fact that this action causes suffering. Or, to choose another example, we justify our belief that it is morally wrong to keep a particular book by arguing that this book is the property of someone else. It is obvious that on this conception of evidence for moral beliefs the beliefs cannot simply be deduced from our evidence. This is so because in contrast to the moral conclusions the evidence has no moral content. We cannot deduce moral propositions from non-moral ones. Or, to put the same point differently, we cannot logically derive what ought to be done from what is. Nevertheless, either background standards or a certain kind of moral sensibility may facilitate our ability to form moral beliefs on the basis of evidence that has no moral content of its own. According to this view, the total evidence (which is shared by us and our moral aliens) *uniquely* determines which moral attitudes, i.e. either belief, disbelief or suspension, are

⁸ However, there may be some room for explaining away this intuition such that our hostile attitude towards the moral alien is a matter of fact rather than what is rationally required. One also might argue that we cannot suspend acting such that there is in practice no room for being agnostic.

rational. This uniqueness thesis is supported by the well-known claim that moral facts supervene on non-moral facts.

This leads to the following *argument from undercutting defeat*. We and our moral aliens share all the evidence that is relevant for the justification of moral beliefs. This is true because we assume that all the evidence for moral beliefs is evidence *without moral content*. By stipulation, this evidence is fully shared between us and our moral aliens. At most one of us can have justified moral beliefs because the relevant evidence is shared and uniquely determines which moral beliefs are justified. Since our epistemic situation is fully disclosed to us, we also know that at most one of us—either the moral aliens or us—has justified moral beliefs. Since our moral aliens are as intelligent and thoughtful as we are and seems to have as internally coherent moral beliefs as we have, we as moral realists have no reason to privilege our own perspective. We therefore must conclude that there is a significant chance that we do not adequately respond to the shared evidence. This undermines our justification for our moral beliefs even if we are in fact assessing the evidence correctly. Hence, we must significantly downgrade our moral confidence.

Note that although this argument is conciliationist in spirit, it does not rely on any specific principle from the epistemology of disagreement.⁹ It just uses general insights into the dynamics of rational revision of beliefs under defeating evidence. The situation above is similar to the Restaurant Case in which I learn that someone whom I reasonably take to be as competent in mental calculation as myself comes to a different solution to a simple mathematical problem.¹⁰ In that case, my initial justification is undermined even if I, in fact, calculated correctly. The above argument is also neutral with respect to epistemic internalism and externalism since both are committed to acknowledge the defeasibility of justification.

So far, I have argued that there is rational pressure on the moral realist to suspend her moral beliefs upon encountering moral aliens, *if the evidence for their moral beliefs is fully shared*. However, we can also construe the epistemic situation in such a way that both parties have very different bodies of *moral* evidence. Here

⁹ However, it is not fully neutral. For example, it is incompatible with Kelly's Right Reason View according to which we are fully rational in remaining steadfast in the face of disagreement if we reasoned correctly (Thomas Kelly, "The Epistemic Significance of Disagreement," *Oxford Studies in Epistemology* 1 (2005): 167–196). Kelly has given up this view in the meantime because he thinks that higher-order evidence has at least some epistemic significance.

¹⁰ See David Christensen, "Epistemology of Disagreement: The Good News," *Philosophical Review* 116 (2007): 187–217.

are two different ways of spelling this out. First, we may assume that moral *intuitions* about particular cases or about moral principles constitute our moral evidence.¹¹ In this case, it seems plausible that the moral alien possesses a body of evidence that is very different from ours since she has different moral intuitions. For example, she may find it intuitively correct that suffering is morally irrelevant. Alternatively, we may assume that moral *perception* constitutes moral evidence.¹² This would mean that people directly observe whether an action is morally right or morally wrong. It then seems plausible that the moral alien would have moral perceptions that are very different from ours for much the same reasons as with respect to intuition. In both cases, the body of moral evidence is not shared between the two parties, even if they learn that they have different moral intuitions or moral perceptions. Why is that? When I learn what someone morally intuits (or perceives) and in this way learn that she intuits and perceives differently, I do not thereby acquire her intuitions or perceptions. Hence, even if I know of the alien's radically different body of moral evidence, I do not thereby share this evidence. We may call evidence that is not shareable through communication "private evidence."¹³

This gives us all that is required to run the *argument from rebutting defeat*: We have our private moral evidence that supports our own moral perspective. We also have reason to believe that the moral aliens have their own private evidence that, since it is radically different from ours, sufficiently supports their alien moral perspective. According to moral realism, both moral perspectives are in conflict with each other. The truth of one excludes the truth of the other. When we realize all of this, we come to know that the moral alien is sufficiently justified in disbelieving what I believe. But then we must accept that there are also strong reasons against our own moral beliefs and this rationally requires us to revise all our moral beliefs.

Taken together, the two arguments robustly support (3): No matter whether we attribute shared morally relevant evidence to the alien or not, the realist seems to be committed to the view that by encountering the moral alien, we acquire evidence that defeats the *prima facie* justification of our moral beliefs in one way

¹¹ Robert Audi, *The Good in the Right: A Theory of Intuition and Normative Value* (Princeton: Princeton University Press, 2005); Michael Huemer, *Ethical Intuitionism* (Houndmills and New York: Palgrave Macmillan, 2005).

¹² See Robert Audi, *Moral Perception* (Princeton: Princeton University Press, 2015).

¹³ See Richard Feldman, "Reasonable Religious Disagreement," in *Philosophers Without Good*, ed. Louise Antony (Oxford: Oxford University Press 2007), 194-214.

or another and thus leads to rational revision or even suspension of judgment. This result is in conflict with (2).

By contrast, moral anti-realists clearly have the resources to explain (2) in a pretty straightforward manner. Emotivists may claim that the speakers' moral judgments express their own affective attitudes. It is obvious that, in light of our own affective attitudes, the moral alien performs badly. Other moral anti-realists understand moral utterances as propositional, but analyze the propositional content of the utterance as dependent on the speaker's attitudes or the moral norms of her community. On these (indexical) relativist views, it is obvious why we should not give in to the moral aliens. When asserting "Actions that cause suffering are morally wrong" we mean to express our belief that *according to our moral code, actions that cause suffering are morally wrong*.¹⁴ We should not revise the belief when we encounter a moral alien who, by asserting "Actions that cause suffering are not morally wrong," expresses the belief that *according to her moral code, actions that cause suffering are not morally wrong*. Both beliefs are compatible and thus there is no reason to revise one's judgment. The two parties are simply talking past each other. Finally, there is the view that our moral assertions genuinely conflict with the alien's moral assertions but that their truth can be assessed relative to different perspectives of assessment (ours and theirs). This (genuine) truth-relativism can nicely accommodate the intuition that there is a genuine disagreement between us and moral aliens.¹⁵ It also explains why we rely on our perspective rather than the alien's when we assess the truth of our moral judgments. According to this perspective, we get it right and they get it wrong. This is exactly, what our intuition (2) claims about the case.

2. The Problem of Moral Aliens: The Realist's Solution

Given the inconsistent quartet I introduced in section I, moral realism can only be maintained if we give up (1), (2) or (3). Since I find (1) and (2) extremely plausible, I will focus my criticism on (3). In this section, I will argue that although (3) looks convincing in the light of the two arguments given above, it is ultimately false for reasons that are independent of one's views concerning epistemological internalism and externalism or concerning the epistemic significance of disagreement. There is simply no way of interpreting the recognition of moral aliens in such a way that it generates defeaters that remove or reduce the *prima facie* justification of our moral beliefs. Encountering moral aliens does not put any

¹⁴ See Max Kölbel, "Indexical Relativism versus Genuine Relativism," *International Journal of Philosophical Studies* 12 (2004): 300-303.

¹⁵ Kölbel, "Indexical Relativism versus Genuine Relativism," 306.

rational pressure on the moral realist to revise her moral beliefs. Or so I will argue in this section.

First, I will discuss the argument from undercutting defeat:

- | | |
|--|----------------------------|
| (1) The moral alien and I have radically conflicting moral beliefs such that at most one of us can be right. | (from moral realism) |
| (2) The moral alien and I base our moral beliefs on fully shared non-moral first order evidence. | (by stipulation) |
| (3) It is impossible that shared non-moral evidence supports conflicting moral beliefs for different agents. | (from moral supervenience) |
| (4) At least one of us holds moral beliefs that are not supported by the shared evidence. | (from 1, 2, 3) |
| (5) If at least one of us holds moral beliefs that are unsupported by the evidence, the likelihood that it is me is sufficiently high. | (from epistemic peerhood) |
| (6) The likelihood of my moral beliefs being unsupported by my first order evidence is sufficiently high. | (from 4, 5) |
| (7) I sufficiently justify (6) by deducing it from the justified premises (1), (2), (3) and (5). | (by assumption) |
| (8) If I am sufficiently justified in believing that the likelihood of my beliefs being unsupported by my first order evidence is sufficiently high, I can no longer rationally use my first order evidence to justify my beliefs. | (undercutting defeat) |
| (9) I can no longer rationally use my first order evidence to justify my moral beliefs. | (from 7, 8) |

I already said a bit about the motivation of the premises used in this argument before. Let me here add only a few remarks. (Ad1) If moral realism is correct, there is substantial disagreement between me and the moral alien such

that at most one of us can be right (if we hold contrary but not contradictory views, both of us may be mistaken). (Ad2) As I understand moral aliens, they share all our non-moral evidence by definition. As I already indicated above, this argument relies on the assumption that all the morally relevant evidence has non-moral content. Together, this motivates (2). (Ad3) In order to be rational, the formation of moral judgments on the basis of non-moral evidence has to respect that moral facts supervene on or are grounded in non-moral facts, i.e. that there cannot be a difference in the moral value/normative status of an act without there being a difference in the non-moral properties of that act. (Ad5) This premise articulates the idea that moral aliens have a significant epistemic weight relative to me (epistemic peers). This is motivated by the stipulation that my alien is approximately as intelligent and thoughtful in her judgments as I am and that her system of moral beliefs is as coherent as mine. (Ad7) This premise makes sure that the premises of the argument from undercutting defeat are *justified*. This is needed to produce a defeater since defeaters have to be justified in order to do their job of removing prima facie justification. I cannot give a full defense of this assumption here. What is crucial is that defeaters play the role of reasons against beliefs; and naked (unjustified) beliefs do not constitute reasons.¹⁶ As we will see below, the fact that all the premises must be justified rather than true is the crucial weak point of this argument. (Ad8) This premise simply expresses the epistemic principle of undercutting defeat.

Although the argument from undercutting defeat does not explicitly rely on any assumptions about the epistemology of disagreement, there is an easy way out for non-conciliationists. They would object either to premise (5) or (8). The proponents of a non-conciliatory or steadfast position typically believe that in evaluating the justificatory status of beliefs in the face of disagreement we should not fully ignore our first order evidence.¹⁷ Accordingly, they may argue that (5) is false if we have in fact correctly reasoned from our non-moral evidence. In this case, the majority of evidence would be on our rather than on the alien's side. Alternatively, proponents of a steadfast position might also argue that (8) is false because suggesting that there is a realistic chance of having reasoned improperly

¹⁶ See William Alston, "Philosophy, Naturalism, and Defeat," in *Naturalism Defeated?*, ed. James K. Beilby (Ithaca and New York: Cornell University Press, 2002), Timothy Loughrist, *Reasons against Belief. A Theory of Epistemic Defeat* (Pro Quest Dissertations & Theses A&I, 2015), and David Alexander, "Unjustified Defeaters," *Erkenntnis* 82 (2017): 891-912, for a defense of the justificatory requirement for epistemic defeaters.

¹⁷ Thomas Kelly, "Peer Disagreement and Higher Order Evidence," in *Social Epistemology: Essential Readings*, eds. Alvin Goldman and Dennis Whitcomb (Oxford: Oxford University Press, 2011), 201.

does not completely screen off the significance of the first order evidence but only reduces it within in the body of total evidence.

I do not take this route since I want to argue that the argument from undercutting defeat is unconvincing no matter what the correct view on the epistemology of disagreement is. For this reason, my argument continues under the less comfortable assumption that conciliationism is the correct view. As I indicated above, the argument from undercutting defeat succeeds only if premise (5) is justified. But can we justify the premise that my peer and I are epistemic peers, i.e. that we are equally competent in reasoning on the basis of the morally relevant evidence? So far, I just assumed that equal intelligence, thoughtfulness and coherence on both sides are sufficient to justify epistemic symmetry. But in fact, we need to establish more than this. In principle, there are three different ways of establishing epistemic peerhood. First, we may use a *track-record argument*. When using this kind of argument, we rely on our comparative assessment of past performances of agents or instruments. If they produced roughly the same ratio of correct results in the past, we infer inductively that they are performing equally well in general. Elo ratings in chess play exactly this role. Secondly, we can justify that two people are peers in a specific domain if there are *good indirect indicators* for them being equally competent in their judgments about the domain. We use this source when we attribute peerhood to people who do mental calculation or report what they see. In these cases, we assume that *normal* people are equally good at mental math or perceptual judgments. As long as there is no further sign of radical inequality, we are justified in believing that agents who appear to be normal are equally competent. Thirdly, one might think that we are justified by default to assume epistemic equality unless we have positive evidence that one of us is superior. Since Henry Sidgwick was the first to endorse this principle explicitly in his argument for suspending judgment in the face of disagreement, I will refer to it as *Sidgwick's principle*.¹⁸

I will now argue that we cannot use any of these routes to justify the epistemic peerhood of moral aliens. First, we cannot use a track record argument to justify the peerhood of moral aliens. This is so because, from our perspective, the moral reasoning of moral aliens looks terribly poor. From our perspective, they have always come to ridiculous moral conclusions when they rely on the same non-moral evidence as we do. Hence, there is no basis for supporting the attribution of equal competence to moral aliens by track record arguments.

¹⁸ See Henry Sidgwick, *The Methods of Ethics* (Indianapolis and Cambridge: Hackett Publishing Co., 7th edition, 1907), 342; Sarah McGrath, "Moral Disagreement and Moral Expertise," *Oxford Studies in Metaethics* 4 (2007): 91-92.

Secondly, what about indirect indicators of epistemic peerness with respect to moral beliefs? We know about the moral alien that she is as intelligent and thoughtful as we are. Moreover, her moral belief system seems to be coherent to the same degree as our moral belief system. Are these good indirect indicators of peerness? I don't think so. On the one hand, intelligence and thoughtfulness are general abilities that are not significantly correlated with highly domain-specific competences like moral reasoning. On the other hand, the degree of coherence among moral conclusions does not systematically reflect the degree to which they respect the evidence. For example, one might be strongly and robustly biased in one's reasoning towards a certain direction. This may lead to mutually coherent conclusions that are completely inadequate responses to one's evidence. We cannot argue that moral aliens are normal moral thinkers by reference to their internal rationality because they are not. Any *prima facie* justified assumption of normality is defeated as soon as we discover that moral aliens have radically different moral beliefs. At least one of us has radically mistaken moral beliefs. Therefore, it is not possible to argue for the epistemic peerness of moral aliens in this way either.

Thirdly, one might use Sidgwick's principle to defend the peerness-assumption concerning moral aliens and thereby motivate our suspension of moral judgment. One might argue as follows: when we encounter a moral alien, we realize that there are two radically different moral perspectives. One of them is ours, the other is the moral alien's. We know that they cannot both be correct, but we have no reason to privilege one of them. Of course, when we rely on our own perspective, the alien's judgments look terrible. But the same is true about us from her perspective. There is simply no perspective accessible to us that permits an impartial assessment of both points of view. Without having any reason to prefer one point of view to the other, we should treat both as equally weighty. For this reason, we should suspend our moral judgment after all. This is the line of reasoning suggested by Sidgwick's principle according to which one ought to treat points of view as epistemically equal unless one has reasons to believe that one is superior to the other. However, this principle licenses verdicts that are intuitively too strong.¹⁹ Consider the following case of an *Epistemic Troublemaker*. Suppose you are giving a public lecture at a foreign university to a very diverse audience. Students sit with faculty and interested laypeople from town. With the exception of the colleague who invited and introduced you, none of the people are known to you. In your talk you defend the proposition that p on a topic in normative ethics. At the end of your talk, someone unknown to you stands up and asserts with a

¹⁹ See also David Christensen, "Disagreement, Question-Begging and Epistemic Self-Criticism," *Philosopher's Imprint* 11 (2011): 15-16; Vavova, "Moral Disagreement and Moral Skepticism."

serious tone “*p* is false.” Without saying anything else she leaves the room. If Sidgwick’s principle were correct, the troublemaker’s intervention would be sufficient to remove justification from your belief that *p*. This is so because you have no reason to regard the unknown opponent as epistemically inferior to you. But this consequence seems absurd. Justification cannot be lost so easily.

On closer inspection, the argument from undercutting defeat cannot be upheld. The proponent of a steadfast position will object to premise (5) or premise (8). Even on a conciliationist view, one must accept that there is no way to justify the peeriness-assumption concerning moral aliens as it is required by premise (5). In conclusion, the argument fails on both accounts. Note that this criticism does not depend on any commitment to either epistemological internalism or externalism. Either view must accept relevantly similar requirements for epistemic defeat.

Does the *argument from rebutting defeat* fare better? The core idea of the argument is the following: we assume that we and our moral aliens possess different but on both sides sufficiently good private moral evidence. As soon as we acknowledge this fact, we have to accept that there is strong evidence for and against our own moral position. However, if there are not only strong reasons in support of our position but also strong reason against it, a significant revision of our initial position is rationally required. Here is a semi-formal version of the *argument from rebutting defeat* that can be generalized to all moral propositions *p*:

- (1) My moral belief that *p* is sufficiently justified. (by assumption)
- (2) I am justified in believing that my moral alien is sufficiently justified in believing that not-*p*. (premise)
- (3) If someone is sufficiently justified in believing that someone else is sufficiently justified in believing not-*p*, then she is sufficiently justified in believing not-*p*. (premise)
- (4) I am sufficiently justified in believing not-*p*. (from 2,3)
- (5) I am sufficiently justified in believing *p* and believing not-*p*. (from 1,4)
- (6) If I am sufficiently justified in believing *p* and believing not-*p*, (rebutting defeat)

believing *p* is no longer sufficiently justified.

- (7) Believing *p* is no longer sufficiently justified. (from 5,6)

In what follows, I will argue that no matter whether we understand justification in an externalist (reliabilist) or internalist (mentalistic) way one of the argument's premises will be false—though not the same. Let me start with the reliabilist understanding of justification. Roughly speaking, on a reliabilist account, beliefs are justified only if they are produced by reliable processes that result predominantly in true beliefs. Given this understanding of epistemic justification, premise (2) cannot be true. When we encounter the moral alien, her moral beliefs appear to be widely mistaken and thus the underlying processes appear to be unreliable from our perspective. Hence, we have no reason to believe that they are sufficiently reliable.

By contrast, what does the epistemic situation look like if we understand epistemic justification along the lines of internalist mentalism? Then, justification does not require objective reliability. Hence, we need not attribute reliable mechanisms of forming moral beliefs to the moral alien in order to attribute justified moral beliefs to her. For having the justified belief that *p* it might be sufficient that it seems or appears (in some non-doxastic sense) true to the believer that *p*. In this sense, we can attribute justified moral beliefs to the moral alien, and we can be justified in doing so. We just need to attribute corresponding moral intuitions or moral perceptions to her. On this account, premise (2) is satisfied. But now premise (3) turns out to be problematic. Here is why: Replace “sufficiently justified” in (3) by “appears true.” Then we get: If it appears true to someone (say A) that it appears true to someone else (say B) that not-*p*, then it appears true to A that not-*p*. But why should that be true? Just by acknowledging the appearance of some proposition's truth to someone else, this proposition need not appear true to me. To ascribe justified beliefs to someone else need not justify these beliefs for me.²⁰ We need a bridging principle to arrive at the required connection. If we

²⁰ Alternatively, one might use Feldman's principle that evidence of evidence for *p* is evidence for *p* itself. However, there are many objections to this principle—in particular, if it is applied intersubjectively. See Feldman, “Reasonable Religious Disagreement,” 151, and the critical discussion in Branden Fitelson, “Evidence of Evidence is not (Necessarily) Evidence,” *Analysis* 72 (2012): 85-88; Juan Comesana and Eyal Tal, “Evidence of Evidence is Evidence (Trivially),” *Analysis* 75 (2015): 557-559; Luca Moretti, “Tal and Comesana on Evidence of Evidence,” *The Reasoner* 10 (2016): 38-39; William Roche, “Evidence of Evidence is Evidence under Screening-

attribute the appearance of p 's truth to the alien and if we also assume that the alien's appearances are reliable indicators of truths about the relevant domain, then it should appear true to us that p . However, we cannot add this further assumption in the case at hand for the simple reason that moral aliens do not appear to us as being reliable about the moral domain. We cannot avoid believing that they massively misrepresent moral reality by their appearances.

4. Conclusion

In this paper, I started out with a problem for moral realism, the problem of moral aliens. On the one hand, moral realists want to claim that we should stick to our guns when we encounter radically different moral views. On the other hand, moral realists seem to be committed to the view that there can be a certain epistemic symmetry between us and our moral aliens that forces us into rational suspension of our moral beliefs. Unless one disputes the very possibility of moral aliens, this poses a severe challenge to the moral realist—a challenge that was articulated by the inconsistent quartet.

On closer scrutiny, it turned out that we cannot make any sense of the idea that the moral aliens should be taken as our epistemic peers. The epistemic asymmetry between us and them is inescapable. Interestingly, my argument does not rely on any meta-epistemological or methodological background assumptions. No matter whether one is an internalist or externalist, a steadfast or a proponent of conciliationism, there is simply no way to argue that encountering a moral alien gives us any reason to revise our moral beliefs. If this is correct, the possibility of meeting moral aliens poses no real challenge to moral realism.²¹

Off," *Episteme* 11 (2014): 119-124, Eyal Tal and Juan Comesana, "Is Evidence of Evidence Evidence?," *Nous* 51 (2017): 95-112.

²¹ **Acknowledgements:** Early versions of this paper were presented at the Conference on Moral Disagreement, organized by the ACU at Rome in March 2018, and at the Concept Epistemology Brownbag at the University of Cologne in May 2018. I am particularly grateful to Robert Audi and Massimo Dell'Utri for their extremely valuable written comments on my paper. Further comments from and discussions with the following colleagues also helped me to work out the final version of this paper: Robert Adams, Mark Alfano, Dominik Balg, Sven Bernecker, Sofia Bokros, Jan Constantin, Roger Crisp, Amy Flowerree, Michael Moore, Jakob Ohlhorst, Francesco Praolini, Paul Silva. I am extremely grateful to all of them.

THE SUBJECT'S PERSPECTIVE OBJECTION TO EXTERNALISM AND WHY IT FAILS

Perry HENDRICKS

ABSTRACT: The subject's perspective objection (SPO) is an objection against externalist theories of justification, warrant, and knowledge. In this article, I show that externalists can accommodate the SPO while remaining externalist. So, even if the SPO is successful, it does not motivate internalism, and the primary motivation for internalism has been lost. After this, I provide an explanation for why so many people find cases that motivate the SPO convincing.

KEYWORDS: subject's perspective objection, externalism, internalism, justification

1. Introduction

The subject's perspective objection (SPO) is commonly used to motivate internalist theories of justification; indeed, Michael Bergmann portrays it as the *main* reason for endorsing internalism.¹ Variations of the SPO have appeared in, for example, Laurence Bonjour,² Keith Lehrer,³ Paul Moser,⁴ and Bruce Russell.⁵ In this article, I explain internalism, externalism, and the SPO. Next, I show that one can accommodate the SPO *while remaining an externalist*, meaning that the SPO doesn't motivate internalism. Therefore, the main motivation for internalism has been lost, and the case for internalism is substantially weakened. After this I show that the SPO rests on a false premise, and hence does not threaten externalism.

¹ See Michael Bergmann, *Justification Without Awareness* (New York: Oxford University Press, 2006).

² Laurence Bonjour, "Externalist Theories of Empirical Knowledge," *Midwest Studies in Philosophy* 5, 1980: 53-73 and Laurence Bonjour *The Structure of Empirical Knowledge* (Cambridge: Harvard University Press, 1985).

³ Keith Lehrer, *Theories of Knowledge* (Nashville: Westview Press 1990) and Keith Lehrer "Proper Function and Systematic Coherence," in *Warrant in Contemporary Epistemology: Essays in Honor of Plantinga's Theory of Knowledge*, ed. Jonathan Kvanvig (Pennsylvania: Rowman and Littlefield, 1996): 25-45.

⁴ Paul Moser, *Empirical Justification* (Dordrecht Holland: D. Reidel, 1985).

⁵ Bruce Russell, "The Problem of Evil and Replies to Some Important Responses," *European Journal of Philosophy of Religion* 10, 3 (2018): 105-131.

Finally, I provide a diagnosis for why people find the cases that motivate the SPO convincing.

2. The Subject's Perspective Objection

In epistemology, internalism and externalism are the dominant positions.⁶ Internalists hold, roughly, that if S is justified in believing p, then S is (actually or potentially) aware of what justifies her belief that p. Externalists deny this: externalists hold that S can be justified in believing that p even if she isn't (actually or potentially) aware of what justifies her belief that p. Many have objected to externalism on the grounds that if S is not (actually or potentially) aware of what justifies her belief then, *even if it meets externalist conditions for justification*, the truth of it will appear accidental to her. However, if S's belief appears accidentally true to her, then she isn't justified in believing it. Hence, externalism is false and we should endorse internalism.

Laurence Bonjour, perhaps the most famous proponent of this style of objection, uses the following story to motivate internalism:

NORMAN: Norman, under certain conditions that usually obtain, is a completely reliable clairvoyant with respect to certain kinds of subject matter. He possesses no evidence or reasons of any kind for or against the general possibility of such a cognitive power, or for or against the thesis that he possesses it. One day Norman comes to believe that the President is in New York City, though he has no evidence either for or against his belief. In fact the belief is true and results from his clairvoyant power under circumstances in which it is completely reliable.⁷

In NORMAN, externalist conditions for justification are met, yet Norman lacks justification. Hence, externalism is false. As Bonjour tells the story, we are supposed to understand Norman as not having positive reasons to doubt that veracity of his belief; he doesn't have a defeater for his belief.⁸ However, just one page later, Bonjour suggests that

[I]t becomes quite difficult to understand what Norman himself thinks is going on. *From his standpoint*, there is apparently no way in which he *could* know the President's whereabouts...Why isn't the mere fact that there is no way, as far as

⁶ Though, they are not exhaustive. Michael Bergmann argues that a position he calls 'mentalism' is neither internalist nor externalist. That said, I will treat the positions as exhaustive in this article. See Michael Bergmann, *Justification Without Awareness* (New York: Oxford University Press, 2006).

⁷ Bonjour, *The Structure*, 41.

⁸ Bonjour, *The Structure*, 40-41.

The Subject's Perspective Objection to Externalism and Why it Fails

he knows, for him to have obtained [information about the president] sufficient reason for classifying this belief as an unfounded hunch and ceasing to accept it?⁹

So, Bonjour asks us to understand Norman's belief as being akin to an *unfounded hunch*. If we are to think of Norman's belief as being akin to an unfounded hunch, then it seems that we are attributing to Norman a psychological property, namely the psychological property of his belief appearing to him as an unfounded hunch.¹⁰

There is textual evidence for interpreting Bonjour in this way: he says that "[f]rom [Norman's] subjective perspective, *it is an accident that the belief is true*."¹¹ In other words, the belief appears to Norman to be accidentally true.¹² Thus, it appears that a (negative) psychological property, that of appearing accidentally true, accompanies Norman's belief. However, if a belief appears accidentally true to a person, then she has a *defeater* for her belief (more on this later), and this explains why NORMAN threatens externalism: externalist conditions for justification obtain yet his belief is defeated and therefore unjustified. But this explicitly conflicts with Bonjour's urging that we should understand everything from Norman's perspective to be just fine; he doesn't want us to understand Norman as having a defeater. It appears, therefore, that Bonjour is inconsistent in his portrayal of Norman. What should we do here? I suggest that we understand Bonjour as saying that Norman does indeed have a defeater. This is because Norman's belief appearing accidentally true to him is what does the work in NORMAN: if Norman's belief doesn't appear accidentally true to him, then it's not at all clear that NORMAN is a counterexample to externalism or motivates internalism.¹³

Another advantage of this interpretation is that it fits well with other, similar objections to externalism, such as Lehrer's Mr. Truetemp case.¹⁴ Mr. Truetemp has, unbeknownst to him, a device implanted in him that produces

⁹ Bonjour, *The Structure*, 42.

¹⁰ An unfounded hunch differs from other beliefs in how it feels: it feels like a hunch, as opposed to a normal belief.

¹¹ Bonjour, *The Structure*, 43.

¹² This interpretation is forced on us by the fact that Bonjour connects the accidental appearance of Norman's belief to his subjective perspective.

¹³ Bonjour says "[h]ow [can external conditions] justify Norman's belief? *From his subjective perspective, it is an accident that the belief is true*. And the suggestion here is that the rationality or justifiability of Norman's belief should be judged from Norman's own perspective." (*The Structure*, 42-43) This strongly suggests that the appearance of accidentality, which is a defeater for Norman, is what does the work in NORMAN.

¹⁴ Lehrer, "Proper Function."

accurate beliefs about the temperature in such a way that externalist conditions for justification are satisfied. However, Lehrer says, Mr. Truetemp “has no idea whether [his belief about the temperature] is correct and *he is totally mystified by the existence of it*,”¹⁵ and hence externalist conditions for justification are insufficient: Mr. Truetemp’s mystification acts as a defeater for his belief, making it unjustified.¹⁶ Lehrer’s basic point appears to be the same as Bonjour’s: if only externalist conditions for justification are met, the subject’s belief will appear accidental, and this defeats her belief, making it unjustified.

Michael Bergmann has usefully summarized the above style of objection as follows:

The Subject’s Perspective Objection (SPO): If [a] the subject holding a belief isn’t aware of what that belief has going for it, then [b] she isn’t aware of how its status is any different from a stray hunch or an arbitrary conviction. [c] From that we may conclude that from her perspective it is an accident that her belief is true. And that implies that it isn’t a justified belief.¹⁷

[a] is just a case in which internalist conditions of justification don’t obtain, [b] is what Bonjour (and Lehrer) assert follows from [a], and [c], thinks Bonjour (and Lehrer), follows from [b]. So, the SPO maps onto what Bonjour is arguing with NORMAN and the inferences he makes. Thus, I will treat the SPO as representative of NORMAN (as well as Mr. Truetemp and other similar cases). The SPO, according to Bergmann, is the main motivation for internalism. So, if it can be shown that it doesn’t actually motivate internalism, then internalism is in trouble. In what follows, I will try to show just this.

3. An Externalist Solution to the SPO

So, the main motivation for internalism about justification is the SPO. In this section, I will argue that the SPO doesn’t support internalism; rather, it merely supports a no-defeaters condition (explained below). As we saw above, part of the SPO, namely [c], is the claim that S’s belief *appears accidentally true*, and this, claims the SPO, implies that the belief is not justified. We may put this as:

ACCIDENT: If S’s belief that p appears accidentally true to S, then she has a defeater for p.

¹⁵ Lehrer, “Proper Function,” 32, emphasis mine.

¹⁶ For other similar examples, see Lehrer, *Theories of Knowledge*, Moser, *Empirical Justification*, and Russell, “The Problem.”

¹⁷ Bergmann, *Justification*, 12.

The Subject's Perspective Objection to Externalism and Why it Fails

ACCIDENT threatens externalist theories of justification. This is because, according to the SPO, if only externalist conditions for justification are satisfied by S, the truth of her beliefs will appear accidental and hence will be defeated. And this means that her beliefs are unjustified.

While ACCIDENT is not incontestable, I will grant its truth here to see where it takes us. Does ACCIDENT conflict with externalism? Not obviously. What is required to avoid the trouble brought about by ACCIDENT is the following thesis:

NO DEFEATERS: S's belief that p is justified only if S does not have any (undefeated) defeaters for her belief that p.

So, the SPO entails ACCIDENT, and ACCIDENT shows that the correct theory of justification must be able to accommodate NO DEFEATERS. Therefore, if the SPO motivates internalism, it must be that in order to satisfy NO DEFEATERS, S must be aware of that which justifies her belief that p (i.e. it must be that only internalism can accommodate NO DEFEATERS). But, of course, S doesn't need to be aware of that which justifies her belief in order to satisfy NO DEFEATERS; to not have a defeater for p, S needs to *lack* a mental state (whatever belief of hers is acting as a defeater for p), not *have* a mental state (i.e. an awareness of that which justifies her belief that p).

It should be clear, then, that externalism is compatible with, can accommodate, NO DEFEATERS. To illustrate this, suppose one is a proper functionalist about justification:

PROPER FUNCTIONALISM: S is justified in her belief that p if and only if (i) S believes p, (ii) p is true, (iii) S's belief that p was produced by properly functioning cognitive faculties that are successfully aimed at producing true beliefs, and (iv) S is situated in an appropriate cognitive environment.¹⁸

For the adherent of PROPER FUNCTIONALISM to accommodate NO DEFEATERS, she need only add the following condition to PROPER FUNCTIONALISM:

(v) S does not have any (undefeated) defeaters for her belief that p.

¹⁸ For statements and defenses of PROPER FUNCTIONALISM, see Alvin Plantinga, *Warrant and Proper Function* (New York: Oxford University Press, 1993), Alvin Plantinga, *Warrant: The Current Debate* (New York: Oxford University Press, 1993), and Alvin Plantinga, *Warranted Christian Belief* (New York: Oxford University Press, 2000), Bergmann, *Justification*, and Kenneth Boyce and Alvin Plantinga, "Proper Functionalism," in *The Continuum Companion to Epistemology*, ed. Andrew Cullison (London: Continuum, 2012): 124-141.

But, of course, (v) doesn't (at least obviously) entail that S is (potentially or actually) aware of that which justifies her belief that p; it merely means that she doesn't have a reason to doubt the reliability of p's source. (Indeed, Bergmann adds this condition to PROPER FUNCTIONALISM.¹⁹) Hence, NO DEFEATERS is compatible with externalism. That is, NO DEFEATERS is *neutral* in respect to the internalism-externalism debate: all the SPO shows is that NO DEFEATERS must be included in the correct theory of justification, but NO DEFEATERS can be added to both internalist and externalist theories of justification. Therefore, the SPO doesn't motivate internalism: the main motivation for internalism is not a motivation for it at all. So, if the SPO really is the main motivation for internalism, then internalism is in bad shape indeed.

4. Objection: The SPO Does Motivate Internalism

One might think that I have moved too quickly here: while NO DEFEATERS is neutral in respect to internalism and externalism, the SPO provides a way to link NO DEFEATERS to internalism. This is because in the SPO [a] entails [b] and [b] entails [c], and hence if S's belief that p doesn't meet internalist conditions for justification, then p will *always* appear accidentally true to S, and hence S will *always* have a defeater for p. Therefore, if NO DEFEATERS is part of an externalist theory of justification, S will *never* satisfy it²⁰ and hence will *never* be justified by externalist standards. So, the only way to be a realist about justification, to hold that we are actually justified in some of our beliefs, is to assume that internalist conditions for justification are met.²¹ Call this THE ARGUMENT. We may put it as follows:

- 1) If [a] the subject holding a belief isn't aware of what that belief has going for it, then [b] she isn't aware of how its status is any different from a stray hunch or an arbitrary conviction. (from the SPO)
- 2) If [b] S isn't aware of how its status is any different from a stray hunch or arbitrary conviction, then [c] from her perspective it is an accident that her belief is true.²² (from the SPO)
- 3) Therefore, if [a] the subject holding a belief isn't aware of what that belief has going for it, then [c] from her perspective it is an accident that

¹⁹ More precisely, he adds a *no believed* defeaters condition (Bergmann, *Justification*, 163-168).

²⁰ Here, I am supposing that internalist conditions for justification are not met; only externalist conditions are.

²¹ This is interesting since externalism is usually cited as preserving commonsense views about justification.

²² For ease of read, I have slightly modified [b] and [c].

The Subject's Perspective Objection to Externalism and Why it Fails

her belief is true. (from (1) and (2))

- 4) If a belief doesn't meet internalist conditions for justification, then [a].
(from the definition of internalism)
- 5) Therefore, if a belief doesn't meet internalist conditions for justification,
then [c]. (from (3) and (4))
- 6) If [c], then the belief has a defeater and is not justified. (from
ACCIDENT and NO DEFEATERS)
- 7) Therefore, any belief that doesn't satisfy internalist conditions for
justification is not justified. (from (5) and (6))

The upshot of THE ARGUMENT is that *even if* the externalist adds NO DEFEATERS to her criteria of justification, she will never be able to satisfy it; that is, if S is not aware of that which justifies her belief that p (i.e. if she doesn't satisfy internalist conditions for justification), then, per the SPO, she will *never* satisfy NO DEFEATERS and hence will never have pure externalist justification.²³ Therefore, since the SPO rules out (pure) externalist justification ever obtaining, it motivates internalism.

4.1 Response: THE ARGUMENT is Unsound

No doubt THE ARGUMENT, if sound, shows that the SPO supports internalism. However, I will show that both premises (1) and (2) of THE ARGUMENT are false.

The fundamental problem with premise (1) is that it only takes into account awareness of positive aspects of beliefs. While the externalist holds that S can be justified even if she is not aware of its positive aspects and their relevance to her belief, she—if she endorses NO DEFEATERS—will also hold that S is not justified if her belief has a(n undefeated) defeater. If a belief of S's has a defeater, then S can recognize that it does and distinguish it from other, undefeated beliefs. This means that S can recognize that arbitrary convictions or stray hunches are epistemically bad and have defeaters, and this enables her to distinguish them from other beliefs of hers that do not have defeaters. So, if S's belief that p meets the externalist conditions for justification and NO DEFEATERS, then she will be able to distinguish p from another belief p* that has a defeater (e.g. because it is a 'stray hunch' or 'arbitrary conviction'). In other words, S can distinguish defeated beliefs (e.g. a stray hunch) from undefeated beliefs, and hence premise (1) is false.

²³ By "pure externalist justification" I mean justification that results from purely externalist conditions, such as PROPER FUNCTIONALISM.

Crucially, *p* appearing accidentally true from *S*'s perspective is a *psychological property*; therefore, in considering whether premise (2) is true, we need to know whether [b] necessitates a certain psychological property. [b], in essence, states that any belief of *S*'s that is in the epistemic class of hunches and arbitrary convictions will have the psychological property of appearing accidental from her perspective. However, once we recognize this, premise (2) seems highly dubious. Consider a 'stray hunch.' One might have a stray hunch that *X* is true, and when she finds out it's true, she will say "I knew it! I told you so!," in which case the truth of *X* didn't appear accidental to *S*.²⁴ (More generally, it is dubious to suppose the psychological state *appearing accidental to S* is *necessarily* connected with *S*'s belief being akin to a stray hunch.) So the truth of a stray hunch need not (necessarily) appear accidental to *S*. But perhaps I have overlooked the fact that it is a *stray* hunch: one might claim that to have a *stray* hunch just is for one to have a hunch that, after turning out true, appears accidental to its subject. If that is how we are to read [b], then the objector is correct. However, if we read [b] that way, then [c] reduces to [b]. Since [c], on this interpretation, is just a restatement of [b], premise (2) is superfluous, and we are left with only the inference from [a] to [b] in premise (1), which we have already seen is false. So, either premise (2) is false, or it reduces to premise (1) which is false. Either way, THE ARGUMENT is unsound.

The upshot of this section is that the SPO is fundamentally flawed. Hence, *even if the SPO entailed internalism*, it would not help internalists since it relies on false premises.

5. Why do so Many Think the SPO is Plausible?

The reason the SPO strikes many as being so plausible, I contend, is that the examples that are usually given to motivate it make use of cognitive faculties that humans are predisposed to regard as dubious. Thus, the deck has been rigged against externalism. The formula of the SPO is this:

FORMULA: (i) *S* has cognitive faculty *F*, (ii) *S* forms a belief by way of *F*, (iii) *F* reliably produces true beliefs, and (iv) *S* has no way of confirming (or never has confirmed) (iii).

Typically, condition (i) of FORMULA refers to a *F* that humans don't have.²⁵ It is then argued that in FORMULA externalist conditions of justification are met,

²⁴ This is because we don't claim to know things that appear accidentally true. Instead, we might say "that was lucky."

²⁵ Norman is a clairvoyant (Bonjour, *The Structure*), Mr. Truetemp has a temperature faculty (Lehrer, "Proper Function"), Mr. Truenorth has a faculty that produces beliefs about what

but S's belief produced by F isn't justified since it is, in some sense, bizarre. And hence externalism is false. However, we can fill out FORMULA in a way that doesn't undermine externalism. Consider:

NORMA: Norma is a regular human being that is dropped off into the wild by her parents when she is a baby. By divine provision, she grows up into adulthood without encountering another living creature. On her 18th birthday, a group of hikers stumble upon Norma. Upon seeing them, her 'theory of mind'²⁶ produces various beliefs about their (the hiker's) mental states (e.g. *that* thing looks concerned, *that* thing is trying to communicate with me, etc.). She is unaware she has a theory of mind, has never tested it or used it before, and doesn't know the typical cognitive faculties that her species has.

I suspect that many will be inclined to attribute Norma justification. But, like NORMAN (Mr. Truetemp, etc.), NORMA is just a filled out version of FORMULA. This shows us that whether FORMULA supports the SPO is contingent on what faculty is instantiated in F: if F isn't a faculty a normal human possesses, it (may) elicit(s) the intuition that S lacks justification. However, if F is a normal human faculty, then it (may) elicit(s) the intuition that S is justified. This shows us that what's doing work in the SPO is *not* the fact that externalist conditions are met, but that the faculty that produces the belief in question isn't typically had by human beings. So, to hold that these cases support the SPO is to affirm a form of justificatory imperialism: it entails that beliefs produced by cognitive faculties that are not had by humans do not have justification. The grounding of the SPO, therefore, appears to be in a prejudice against non-human-like cognitive faculties. So, those who are dubious about such a human-centered view of justification have further reason to reject the SPO.²⁷

direction one is facing (Russell, "The Problem"), and so on.

²⁶ Theory of mind is a cognitive faculty responsible for producing beliefs about the mental states of other creatures. See e.g. chapter 1 of Justin Barrett, *Why Would Anyone Believe in God?* (Maryland: AltaMira Press, 2004), chapters 1-3 of Justin Barrett, *Cognitive Science, Religion, and Theology: From Human Minds to Divine Minds*. (Pennsylvania: Templeton Press, 2011), and Justin Barrett, *Born Believers: The Science of Children's Religious Belief*. (New York: Free Press, 2012), N. Knight, P. Sousa, J. Barrett, and S. Atran, "Children's attributions of beliefs to humans and God," *Cognitive Science* 28 (2004): 117-126, R.A. Richert and P.L. Harris, "Dualism revisited: Body vs. mind vs. soul," *Journal of Cognition and Culture* 8 (2008): 99-115, and Adam Waytz, Kurt Gray, Nicholas Epley, and Daniel M. Wegner "Causes and consequences of mind perception," *Trends in Cognitive Sciences* 8 (2010): 383-388.

²⁷ For comments on this article, thanks to Michael Bergmann. And thanks especially to G.L.G.—Colin Patrick Mitchell—for particularly insightful comments.

HUSSERLIAN EIDETIC VARIATION AND OBJECTUAL UNDERSTANDING AS A BASIS FOR AN EPISTEMOLOGY OF ESSENCE

Robert MICHELS

ABSTRACT: Vaidya has recently argued that while Husserl's method for acquiring knowledge of essence through use of our imagination is subject to a vicious epistemic circle, we can still use the method to successfully attain objectual understanding of essence. In this paper, I argue that the Husserlian objectual understanding-based epistemology envisaged by Vaidya suffers from a similar epistemic circularity as its knowledge-based foil. I argue that there is a straight-forward solution to this problem, but then raise three serious problems for an amended version of Vaidya's proposal and any similar Husserlian epistemology of essence. The paper closes with some general reflections on applying the Husserlian method to the contemporary notion of essence and on the idea of refocusing the epistemology of essence on understanding instead of knowledge.

KEYWORDS: epistemology of essence, eidetic variation, understanding, essence, Husserl, Vaidya

1. Essence and the Husserlian Approach to the Epistemology of Essence

The notion of essence has recently seen a renaissance in philosophy. Essentialism has first come to prominence again in the context of the discussion about metaphysical modality which followed important advances in quantified modal logic in the 1970s and 1980s. The standard view emerging from the early debate simply identifies essentiality and metaphysical necessity, so that the essential properties of an object are just those for which it is metaphysically necessary that the object has them, if it exists.¹ More recently, philosophers such as Fine, Correia,

¹ See e.g. Saul Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980), Penelope Mackie, *How Things Might Have Been* (Oxford: Clarendon Press, 2006), David Wiggins, "The *De Re* 'Must': a Note on the Logical Form of Essentialist Claims," in *Truth and Meaning*, eds. Gareth Evans and John McDowell (Oxford: Oxford University Press, 1976), 285–312.

and Lowe² have embraced the notion of essentiality, but rejected the idea that it is identical or even definable in terms of metaphysical necessity. This development immediately raises questions about the epistemology of essence. While there are several well-developed approaches to the epistemology of metaphysical modality, such as the conceivability-based approach,³ it is not clear in how far we can depend on these approaches in order to explain knowledge of essence once we have distinguished this notion from that of metaphysical necessity. The epistemology of essence still offers many avenues for research.⁴

A historically important approach to the epistemology of essence which was specifically developed with this notion rather than the notion of metaphysical necessity in mind, is due to Husserl. It is the method of *eidetic seeing* or *essential seeing* ('Wesenserschauung'). Husserl's main idea is that knowledge of essence can be gained through a specific kind of intuition, the mentioned eidetic seeing, which we arrive at by starting from what Husserl calls 'intuition of something individual,'⁵ which is an experience of a particular object. At the core of the method is a process he calls *free variation*. Roughly, this process is supposed to

² See Kit Fine, "Essence and Modality," *Philosophical Perspectives* 8 (1994): 1–16, Kit Fine, "Senses of Essence," in *Modality, Morality, and Belief. Essays in Honor of Ruth Barcan Marcus*, ed. Walter Sinnott-Armstrong (Cambridge: Cambridge University Press, 1995): 53–73, Fabrice Correia, "Generic Essence, Objectual Essence, and Modality," *Noûs* 40, 4 (2006): 753–67, Fabrice Correia, "On the Reduction of Necessity to Essence," *Philosophy and Phenomenological Research* 84, 3 (2012): 639–53, and E. J. Lowe, "Two Notions of Being: Entity and Essence," *Royal Institute of Philosophy Supplements* 62 (2008): 23–48.

³ See e.g. Stephen Yablo, "Is conceivability a guide to possibility?" *Philosophy and Phenomenological Research* 53, 1 (1993): 1–42, David J. Chalmers, "Does Conceivability Entail Possibility?" in *Conceivability and Possibility*, eds. Tamar Szabó Gendler and John Hawthorne (Oxford: Oxford University Press, 2002), 145–200. For an overview of this and other approaches, see Anand Vaidya, "The Epistemology of Modality," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (The Metaphysics Research Lab. CSLI Stanford, 2017). <https://plato.stanford.edu/archives/win2017/entries/modality-epistemology/>.

⁴ Contemporary works in the epistemology of essence include Bob Hale, *Necessary Beings: An Essay on Ontology, Modality, and the Relations Between Them* (Oxford: Oxford University Press, 2013), ch. 11 and Tuomas Tahko, "The Epistemology of Essence," in *Ontology, Modality, Mind: Themes from the Metaphysics of E. J. Lowe*, eds. Alexander Carruth and S. C. Gibb and John Heil (Oxford: Oxford University Press), 93–110, Tuomas E. Tahko, "Empirically-Informed Modal Rationalism," in *Modal Epistemology After Rationalism*, eds. Robert William Fischer and Felipe Leon (Cham: Synthese Library, 2017): 29–45, which build on E. J. Lowe, "What is the Source of Our Knowledge of Modal Truths?" *Mind* 121, 484 (2012): 919–950.

⁵ Edmund Husserl, *Ideas Pertaining to a Pure Phenomenology and to a Phenomenological Philosophy. First book. General Introduction to a Pure Phenomenology* (The Hague: Martinus Nijhoff Publishers, 1983): 8. Original German term: 'individuelle Anschauung.'

work as follows.⁶ First, we have an individual intuition of something which we then take as an example based on which we produce variations of it through our imagination. These variations have to fulfil two conditions. First, they have to be ‘free,’ in the sense that they may randomly deviate from the original experience or intuition, as long as they are still ‘concretely similar’⁷ to it. Second, the number and sequence of variations also needs to be arbitrary, the idea being that it should not matter whether the process of imagining variations is aborted at a particular point or whether it could, at least theoretically, be extended infinitely.⁸ The purpose of the process is to isolate that which is invariant through all variations. This ‘invariable *what*’⁹ is then the essence of the experienced or imagined kind of object, that “without which the object cannot be intuitively imagined as such.”¹⁰

There is of course much more to be said about this method considered in its historical context and in relation to Husserl’s own conception of essence.¹¹ The focus of this paper however is not on exegetical questions, but rather on a different question which we might put as follows: Can a variant of Husserl’s method of essential seeing serve as the basis of an epistemology of essence, assuming the contemporary understanding of the latter notion developed by Fine, Correia, Lowe, and others? This question has been taken up in a recent paper by Vaidya.¹² My main aim in this paper is to assess (and ultimately criticise) Vaidya’s answer to a particular variant of the question which focuses on understanding instead of knowledge as its epistemological target notion.

Vaidya answers the question in the negative for the standard, knowledge-based approach to the epistemology of essence. In this paper, I will take a condensed presentation of Vaidya’s argument for this conclusion as my starting point (section 2). In addition to this negative conclusion, Vaidya argues that our

⁶ Husserl describes the process in detail in § 87 “The method of essential seeing” in Edmund Husserl, *Experience and Judgment. Investigations in a Genealogy of Logic*, ed. Ludwig Landgrebe (London: Routledge & Kegan Paul, 1973): 340–348. See also section II of Daniele De Santis, “Phenomenological Kaleidoscope: Remarks on the Husserlian Method of Eidetic Variation,” in *The New Yearbook for Phenomenology and Phenomenological Philosophy* XI, eds. Burt Hopkins and John Drummond (London: Routledge, 2011): 20–23.

⁷ Husserl, *Experience and Judgment*, 341.

⁸ Husserl, *Experience and Judgment*, 342.

⁹ Husserl, *Experience and Judgment*, 341.

¹⁰ *Ibid.*

¹¹ See in particular Husserl’s unpublished notes on eidetic variation from his Nachlass collected in Edmund Husserl, *Zur Lehre vom Wesen und zur Methode der eidetischen Variation. Texte aus dem Nachlass (1891–1935)*. Husserliana Vol. XLI (Dordrecht: Springer, 2012).

¹² Anand Jayprakash Vaidya. “Understanding and Essence,” *Philosophia* 38, 4 (2010): 811–33.

main question can be answered positively, if we refocus the epistemology of modality on a different state, namely that of understanding. This second, positive argument from Vaidya's paper is presented in section 3, but then criticized in section 4. My main criticism will be that his positive proposal falls victim to an epistemic circularity very similar to the one which Vaidya identified in his critique of the knowledge-based variant of a Husserlian epistemology of essence. I will consider a response to this criticism and then present three arguments against the resulting approach and similar Husserlian approaches in section 5. In section 6, I close with a discussion of the prospects for other applications of Husserl's method in the contemporary epistemology of essence and for Vaidya's idea of refocusing the debate on understanding instead of knowledge.

2. Vaidya's Negative Argument

Vaidya argues that Husserl's method of free variation, which he calls *Variation-in-Imagination*, 'VIM' for short, cannot produce knowledge of essence.¹³ Vaidya's argument for the latter claim is based on three necessary conditions for attaining knowledge of essence. The first is *Necessity-of-object-preservation*, **NOP**:

NOP In order for an imaginative process, such as VIM, to yield a judgment about whether *P* is an essential property of *o* through property variation on *o* via the construction of a set of scenarios *S*₁...*S*_{*n*}, it must be the case that *o* is preserved in the transition from each *S*_{*i*} to *S*_{*k*}.¹⁴

In other words, for **VIM** to successfully produce in us knowledge of the essence of *o*, the scenarios, or in Husserl's term, variations, which we run through,

¹³ See Vaidya, "Understanding and Essence," 820.

¹⁴ Vaidya, "Understanding and Essence," 821. Note that in accepting this condition as a constraint on VIM, Vaidya departs from Husserl, who insists that "variation [in imagination] depends precisely on this: that we drop the identity of the individual and change it imaginatively into another possible individual" (Husserl, *Experience and Judgment*, 347–8). That Husserl's notion of essence admits of no essential truths (in the sense of eidetic law statements) about particular individuals is also explained in Rochus Sowa, "Essences and Eidetic Laws in Edmund Husserl's Descriptive Eidetics," *The New Yearbook for Phenomenology and Phenomenological Philosophy* VII, eds. Burt Hopkins and Steven Crowell (London: Routledge, 2007): 77–108. Note however that Vaidya's acceptance of **NOP** matches the predominant focus on individual essence which one e.g. finds in earlier contributions to the contemporary discussion of essence, such as Fine, "Essence and Modality" and Fine, "Senses of Essence." Later works broadened the focus by including a notion of generic essence (see e.g. Fabrice Correia, "Generic Essence, Objectual Essence, and Modality," *Noûs* 40, 4 (2006): 753–67, Kit Fine, "Unified Foundations for Essence and Ground," *Journal of the American Philosophical Association* 1, 2 (2015): 296–311), which is however also distinct from Husserl's notion.

have to be such that the object o stays the same. It is only the properties of the object which may be varied when passing from one such scenario to another; the object itself cannot be.

The second necessary condition for attaining knowledge of essence through **VIM** is:

NAC If a **VIM** yields knowledge of the essence of o for a subject A , then it cannot be the case that the preservation of o across $S_1...S_n$ is *accidental*.¹⁵

That the objects o is preserved through $S_1...S_n$ just means that o has to be the same object throughout $S_1...S_n$. In other words, **NAC** tells us that for an instance of **VIM** to yield knowledge of essence, it has to *non-accidentally* confirm to **NOP**. This second assumption is modelled on a general assumption about knowledge which is defended by several contemporary epistemologists, namely that it is not merely a matter of (a problematic kind of) epistemic luck that the subject has gained the relevant true belief.¹⁶ The idea captured by **NAC** is accordingly that for an instance of **VIM** to succeed, **NOP** has to be non-accidentally satisfied.

The third and last necessary condition is:

CCC In order for a subject to construct a scenario S_i involving an object o , the subject must consciously choose which properties o is to have from a set of properties I_i , which the subject has knowledge of.¹⁷

CCC's role in the argument is to make clear that it is the subject involved in **VIM** who actively chooses which properties to vary when passing from one scenario to the next.

Briefly stated, the problem Vaidya raises for the Husserlian approach with respect to these three conditions is the following: For **VIM** to reliably yield knowledge of essence, the subject applying the method must consider only variations which involve o (by **NOP**). It furthermore needs to be the case that the relevant true belief generated by **VIM** is not brought about accidentally, i.e. in a manner conducive to an inadmissible case of epistemic luck (by **NAC**). The problem with the process then arises since the preservation of o across the relevant scenarios is a matter of the preservation of o 's essential properties. Since it is the subject involved in **VIM** which has to actively choose which properties to vary when imagining a new scenario (by **CCC**), what is needed to preclude the sort of

¹⁵ Vaidya, "Understanding and Essence," 822.

¹⁶ See e.g. Peter Unger, "An Analysis of Factual Knowledge," *Journal of Philosophy* 65, 6 (1968): 157–170, Duncan Pritchard, *Epistemic Luck* (Oxford: Oxford University Press, 2005). The notion will be discussed in a bit more detail later in the paper.

¹⁷ Vaidya, "Understanding and Essence," 822.

accidentality which is ruled out by **NAC** is a guarantee that the subject will vary only *o*'s accidental properties when going through an instance of **VIM**, leaving its essential properties unchanged.

Vaidya's point is that such a guarantee could only be given if the subject knew beforehand which properties these were. We can try to make this more precise by spelling out the non-accidentality requirement imposed by **NAC** in terms of a set of relevant possible worlds: For **NAC** to be satisfied, it has to be the case that in every relevant possible world in which the subject creates a scenario, the scenario is created by subtracting a non-essential property of *o*, resulting in a scenario centred on *o* and not on some distinct object *o'* which differs from *o* in its essential properties. Since according to **CCC**, the subject creates the scenario by conscious choice, this choice must be backed by a belief that the subtracted property is non-essential to *o*. Assuming that this backing-belief is present in all relevant relevant possible worlds, it follows that this belief of the subject satisfies a modal anti-luck luck condition for knowledge. While this does not logically entail that the subject knows that the relevant property is non-essential to *o*, anti-luck conditions are after all necessary, not sufficient for knowledge, we can, due to a lack of plausible alternative explanations, abductively infer that the subject has to have this piece of knowledge about *o*'s essence. Granting this inference, we can conclude that the subject has to have knowledge of the essence of *o* in order to gain knowledge of the essence of *o* via **VIM**, resulting in a vicious epistemic circle.¹⁸ The, in my opinion correct, conclusion of the negative part of Vaidya's argument is that Husserl's method of free variation is not a suitable means to gain knowledge of essence.

¹⁸ See Vaidya, "Understanding and Essence," 823 for a more detailed presentation of the argument. Note that similar points have been raised by many others authors, including e.g. Peter H. Spader, "Phenomenology and the Claiming of Essential Knowledge," *Husserl Studies* 11, 3 (1994): 179 and Peter Simons, "Experience and Judgment: Investigations in A Genealogy of Logic, by Edmund Husserl," *Journal of the British Society of Phenomenology* 7, 1 (1976): 61–65. A range of similar objections is also critically discussed in David Kasmier, "A Defense of Husserl's Method of Free Variation," in *Epistemology, Archaeology, Ethics. Current Investigations of Husserl's Corpus*, eds. Pol Vandavelde and Sebastian Luft (London: Continuum, 2010): 21–40.

3. Vaidya's Positive Proposal: VIM as a Means to Gain Objectual Understanding of Essence

3.1 Objectual Understanding of Essence

Vaidya argues that **VIM** still has a use in the epistemology of essence: It may not deliver *knowledge*, but it can instead be used to attain *objectual understanding* of essences. Some epistemologists have argued that understanding (including both its objectual and other varieties) is at least as important a subject of epistemological investigation as is knowledge.¹⁹ Vaidya's positive proposal hence promises to give us a new epistemic foundation for the epistemology of essence and potentially also the epistemology of modality more generally.²⁰

There is no universal agreement about what understanding is in the contemporary literature. Among the accounts on offer, Vaidya in particular relies on Kvanvig's. Crucially, this account is based on the idea that 'understanding is not a species of knowledge.'²¹ According to Kvanvig's account, understanding does however share, with an important qualification which will be discussed shortly, an important property of knowledge, namely its factivity. There is a near consensus that knowledge is *factive*, i.e. that for every p , if someone knows that p , then p .²² While the consensus about the factivity of understanding, i.e. about whether the beliefs which an epistemic agent holds when understanding a subject have to be true, is not as strong as in case of knowledge, that some form of factivity holds is still the standard view.²³ Kvanvig in particular points out that while 'understanding' is sometimes used non-factively in order to hedge claims which appear too strong, as e.g. in "My understanding is that you weren't home till after midnight," he still holds that the notion of understanding which is of interest to epistemologists is factive.²⁴

Kvanvig accepts that *propositional understanding*, i.e. the sort of understanding at issue e.g. in "Claude understands that his internet connection is

¹⁹ See Stephen R. Grimm, "The Value of Understanding," *Philosophy Compass* 7, 2 (2012): 103–117.

²⁰ See Vaidya, "Understanding and Essence," 310–312.

²¹ Jonathan L. Kvanvig, *The Value of Knowledge and the Pursuit of Understanding* (Cambridge: Cambridge University Press, 2003), 196.

²² See however Allan Hazlett, "The Myth of Factive Verbs," *Philosophy and Phenomenological Research* 80, 3 (2010): 497–522.

²³ Some non-factive cases of understanding are e.g. admitted in Linda Zagzebski, "Recovering Understanding," *Knowledge, Truth, and Duty: Essays on Epistemic Justification, Responsibility, and Virtue*, ed. Matthias Steup (Oxford: Oxford University Press, 2001).

²⁴ See Kvanvig, *Value of Knowledge*, 191.

slow,” is factive. However, he explicitly relaxes this requirement with respect to *objectual understanding*, understanding of a more complex subject matter, e.g. as in “Claude understands quantum mechanics.”: If someone has some false beliefs about a subject matter, then “we can ascribe understanding based on the rest of the information grasped that is true and contains no falsehoods,”²⁵ as long as these “false beliefs concern matters that are peripheral rather than central to the subject matter in question.”²⁶ Since factivity will play a crucial role in the argument of the next section, I will come back to this point.

Besides the similarity concerning factivity, there are two crucial differences between understanding and knowledge according to Kvanvig’s account. First, “understanding requires, and knowledge does not, an internal grasping or appreciation of how the various elements in a body of information are related to each other in terms of explanatory, logical, probabilistic, and other kinds of relations that coherentists have thought constitutive of justification.”²⁷ To illustrate this, consider Zeno, who has absolutely no idea about quantum mechanics, but has adopted a large set of true beliefs about the theory, including e.g. that the development over time of a quantum system is governed by Schrödinger’s equation, because his physicist friend told him. Zeno can correctly be described as knowing the corresponding facts about quantum mechanics, even though it is clear that he completely lacks the grasping of the intrinsic structure of quantum mechanics which would be required for him to understand quantum mechanics.

Second, unlike knowledge, understanding is compatible with epistemic luck.²⁸ Imagine for example that most books about politics were factually inaccurate, save for the one excellent book which Xenia picked up randomly to learn about politics. The true beliefs about politics which Xenia acquires by reading the book would not constitute knowledge, since they are a product of epistemic luck. Her beliefs are, so to say, not modally stable enough to qualify as knowledge: She was very lucky to have picked the one factually accurate book and could easily have picked one of the many factually inaccurate ones instead. However, in the same scenario, Xenia could still acquire objectual understanding of politics. The presence of epistemic luck would not diminish Xenia’s cognitive achievement of having developed an understanding of politics which manifests itself through her true beliefs about this subject matter.

²⁵ Kvanvig, *Value of Knowledge*, 201.

²⁶ *Ibid.*

²⁷ Kvanvig, *Value of Knowledge*, 192–3.

²⁸ See Kvanvig, *Value of Knowledge*, 199.

3.2 Vaidya's Objectual Understanding-Based Proposal

It is the second difference to knowledge which holds the key to Vaidya's positive proposal, i.e. its compatibility with epistemic luck. Vaidya's idea is to drop both **NAC** and **NOP** as necessary conditions on the success of **VIM** in order to repurpose it as a method for gaining objectual understanding of essence. Dropping **NAC** ensures that the epistemic circularity which affects the method in the knowledge-case no longer arises: The requirement of a guarantee that **NOP** is non-contingently satisfied drops out, which means that we no longer need to assume that a subject has to already have knowledge of the relevant essence.

If the epistemic circularity is already taken care of, why also drop **NOP**? Vaidya gives two reasons: First, the objectual understanding-based version of **VIM** may involve scenarios not centred on the object whose essence the subject aims to discover. According to Vaidya, such scenarios can contribute to the success of **VIM** since they allow the subject to "see how changing a property destroys the object" and "to comprehend why the property in question is essential."²⁹

Second, Vaidya argues that **NOP** is an aboutness condition, which means that it ensures that **VIM** produces objectual understanding of the essence of the right object. This makes **NOP** an implicit second anti-luck condition, since it ensures that the true beliefs which a subject gains through **VIM** about *o*'s essence are not merely by accident about *that* object, rather than another one.³⁰

This second argument for rejecting **NOP** is, to my mind, unconvincing. **NAC** basically says that **NOP** has to be satisfied non-contingently, so if **NOP** itself already has a non-contingency condition built in, why doesn't this render **NAC** redundant? I will not go deeper into this question here for the simple reason that the argument of the next section will settle the status of **NOP** within Vaidya's proposal.

With the epistemic circularity taken care of, the following picture emerges: **VIM** is a method for acquiring objectual understanding of essence. Objectual understanding of essence is a state which i) involves true beliefs about essence, as it is factive, ii) is not subject to a version of **NAC**, as it is compatible with epistemic luck in acquiring these true beliefs, and iii) consists in a cognitively internal grasping of certain relations holding between different aspects of the essence of the object.³¹ Having described Vaidya's positive proposal for an epistemology of essence, I will now criticize it.

²⁹ Vaidya, "Understanding and Essence," 824.

³⁰ See Vaidya, "Understanding and Essence," 826.

³¹ There are several questions about this proposal which one may ask. It does for example not address the metaphysical question of what the object of understanding is when one objectually

4. Another Epistemic Circularity

Vaidya's positive proposal includes the idea that scenarios which violate **NOP** can, and do, play a role in instances of **VIM** which lead to the acquisition of objectual understanding of essence. Here is how Vaidya explains their contribution:

Any scenario that does not contain *o* is a scenario that plays a role in the subject arriving at an objectual understanding of the essence of *o*. In a scenario where some *o** [i.e. an object distinct from *o*] is present the subject can see how changing a property destroys the object. This accidental insight into the object being changed allows the subject to comprehend why the property in question is essential.³²

The general idea captured by this quote is that **NOP**-violating scenarios can contribute to one's gaining objectual understanding of essence by allowing one to see under which conditions the relevant object ceases to be itself. My focus for now will be on instances of **VIM** which involve such scenarios.

It is clear from Vaidya's negative argument that in a scenario of this sort, the subject would have to subtract one of the object *o*'s essential properties. If it tried for example to gain objectual understanding of Socrates's essence, it might, let us assume, do this by subtracting his property of being human. With this kept in mind, let us come back to Vaidya's claim that a **NOP**-violating scenario can contribute to the success of an instance of **VIM** by allowing the subject to see that *o* has ceased to be itself. For the point I am about to make it is important that objectual understanding is factive with respect to the relevant beliefs, i.e. in this case the belief that Socrates is essentially human. As I have pointed out in the previous section, Kvanvig's account of objectual understanding allows for someone to objectually understand a subject matter, even if they have some false beliefs about it, as long as these false beliefs concern only matters peripheral to the subject. Since the following argument requires factivity, I will now discuss to which extent this assumption can be made in the case of objectual understanding of essence, focusing in particular on whether Kvanvig's exception for peripheral beliefs applies in this context.

understands an essence. Kvanvig and others often talk of a subject or subject matter as the object of objectual understanding (see e.g. Kvanvig, *Value of Knowledge*, 197). This view for instance does not seem to square well with the traditional characterization of an object's essence in terms of its essential properties since it is unclear what the subject matter could be in this case. Further questions arise e.g. about the relation between the notion of judgment Vaidya relies on and those of belief and knowledge.

³² Vaidya, "Understanding and Essence," 824.

Granting Kvanvig's exception, which peripheral matters could a subject have false beliefs about without thereby undermining that it objectually understands an entity's essence? Since the essence of a subject matter concerns just those of its features which it has to have for it to be that subject matter,³³ it is hard to see which beliefs about an entity's essence could count as peripheral, rather than central to it. It seems that the only plausible way to classify some beliefs about an entity's essence as peripheral is to accept Fine's distinctions between constitutive/consequential essence (intuitively, those parts of an entity's essence which are directly definitive of it/those parts which aren't, but rather are had essentially in virtue of other parts of its essence) and immediate/mediate essence (intuitively, the mediate essence of an entity also contains the essences of all entities on which it ontologically depends, whereas its immediate essence doesn't) and to identify the peripheral part of its essence with the union of its consequential and mediate parts.³⁴ According to this way of drawing the distinction, at least the intersection of the immediate and constitutive parts of an entity's essence are central in Kvanvig's sense, which means that a subject *must* have true beliefs about them in order to objectively understand the entity's essence. Accordingly, understanding of immediate constitutive essence is indeed factive. In the following, my focus will first be exclusively on such beliefs, i.e. on beliefs about the immediate constitutive essence of entities. I will argue that an epistemic circularity similar to the one exploited by Vaidya in his negative argument will prevent subjects from acquiring understanding of immediate constitutive essence. After making this point, I will argue that this conclusion generalizes to essence in general. For the sake of simplicity, I will use 'essence' as a synonym for 'immediate constitutive essence' throughout the rest of this section, unless explicitly specified otherwise.

³³ As per Aristotle's original characterization of essence as the 'what it is to be'. See his *Metaphysics* Z.4 and all contemporary accounts departing from it, including Fine's, Correia's, and Lowe's.

³⁴ See Fine, "Senses of Essence," sections 3 and 5 for the distinctions and see Kathrin Koslicki, "Varieties of Ontological Dependence," *Metaphysical Grounding: Understanding the Structure of Reality*, eds. Fabrice Correia and Benjamin Schnieder (Cambridge: Cambridge University Press, 2012), 190–195, Eileen S. Nutting and Ben Caplan and Chris Tillman, "Constitutive Essence and Partial Grounding," *Inquiry* 61, 2 (2018): 137–161, and Justin Zylstra, "Constitutive and Consequentialist Essence," *Thought* 8, 3 (2019): 190–199 for discussions of the first of the two distinctions. Note that this identification requires that the consequential parts of the relevant essence are merely consequential, i.e. not also constitutive and that it may be inadmissible for entities whose constitutive essence is essentially inferential, such as e.g. logical concepts like conjunction.

Let us for the sake of simplicity assume that a subject is going through an instance of **VIM** involving only one **NOP**-violating scenario, a scenario in which Socrates is not human and let us furthermore assume that it is part of Socrates's essence that he is human. The crucial point concerning this scenario for which I am going to argue now is that since only true beliefs about the essence of Socrates can constitute objectual understanding, the subject would have to recognize that this scenario does not involve Socrates in order to arrive at a (true) belief which can constitute understanding of Socrates's essence.

To see this point consider what would happen if it did not. In that case, some scenario or some scenarios produced by a subject going through **VIM** would not involve Socrates, but instead a non-human who resembles him in other respects. Now given the following success condition on **VIM** formulated by Husserl, it would follow that the subject acquires a false belief about Socrates's essence: To gain an insight into the essence of an object via an instance of **VIM**, the subject needs to grasp, as Husserl writes, the multiplicity of all the scenarios it has produced. Only by doing that can the subject isolate those features which stay constant throughout all imagined variations of the object and so ultimately gain insight into the object's essence.³⁵ The condition requires the subject to grasp all the scenarios which of course includes any scenario(s) involving the non-human Socrates-like creature. The presence of at least one such scenario would therefore produce a false belief about the essence of Socrates in the subject. To generalize from the example, any instance of **VIM** which involves a **NOP**-violating scenario will produce a false belief in the subject.

Since objectual understanding of essence is factive, a subject cannot acquire objectual understanding of an object's essence via an instance of **VIM** if that instance involves at least one **NOP**-violating scenario. Based on this point, we can, just as in Vaidya's negative argument, abductively infer that a subject would have to recognize the **NOP**-violating scenarios involved in an instance of **VIM** as being such in order to avoid acquiring a false belief about the essence of an entity to finally gain understanding of it. The justification for this abductive inference, just as in Vaidya's argument, is that there is no other plausible explanation for what could enable a subject to successfully apply **VIM** to acquire understanding of the essence of an entity.

³⁵ As Husserl puts it, "[o]nly if we retain in grasp the things imagined earlier, as a multiplicity in an open process, and only if we look toward the congruent and the purely identical, do we attain the *eidos* [i.e. the ideal form which constitutes the essence]" (Husserl, *Experience and Judgment*, 343).

So we face an epistemic circularity again: Only a subject which already has knowledge of the essence of the relevant object, or at least holds true beliefs about the essence of the relevant object as part of its understanding of that essence, can recognize that it has produced a **NOP**-violating scenario. Without this recognition on the side of the subject, the relevant instance of **VIM** will lead it to acquire false beliefs about the essence of the relevant object, leading to a violation of the factivity requirement. Hence, the subject could only successfully gain understanding of the essence of an object via a **NOP**-violating instance of **VIM**, if it already had the true beliefs it is supposed to acquire to gain this understanding. This shows that pace Vaidya, **NOP**-violating scenarios cannot contribute to the success of applications of **VIM**.

Now let us lift the restriction to immediate constitutive essence which was upheld throughout the preceding paragraphs. If **NOP**-violating instances of **VIM** do not allow a subject to acquire objectual understanding of immediate constitutive essence, what does this tell us about objectual understanding of essence in general? Recall that according to Kvanvig, someone can objectually understand a subject even though they hold false beliefs about the subject, as long as those false beliefs concern only peripheral matters, but not matters central to the subject.³⁶ Given the plausible assumption that the matters which are central to an entity's essence are the immediate constitutive parts of its essence, this entails that a subject cannot objectually understand an entity's essence *at all* if it holds false beliefs about that essence's immediate constitutive parts. So for a subject to gain *any sort of* objectual understanding of essence, it has to have only true beliefs about the constitutive immediate parts of this essence. Since I have just (abductively) argued that subjects cannot plausibly satisfy this condition in the context of a **NOP**-violating instance of **VIM**, the conclusion that instances of **VIM** which involve **NOP**-violating scenarios do not allow them to acquire objectual understanding of essence generalizes to essence *tout court*, i.e. the broader notion of essence which also includes consequential and mediate essence.

An immediate consequence of the conclusion of the preceding argument is that a subject can only successfully acquire objectual understanding of an object's essence via an instance of **VIM**, if this instance involves no **NOP**-violating scenarios. But this just means that, pace Vaidya, **NOP** has to be part of the positive proposal. Does re-introducing this principle finally give us a workable understanding-based epistemology of essence?

³⁶ See again Kvanvig, *Value of Knowledge*, 201.

5. Can Vaidya's Proposal be Fixed By Re-Introducing NOP?

Can one simply reintroduce **NOP** as a component of Vaidya's proposal? If Vaidya is right to think of **NOP** as an anti-luck condition, then one cannot consistently do so: Objectual understanding is by definition compatible with epistemic luck, so an epistemology of essence centred on this notion cannot involve an anti-luck condition. There are two ways around this problem: Pace Vaidya, one might just deny that **NOP** is an anti-luck condition. This would immediately dissolve the problem and **NOP** could consistently be retained as a condition on the success of **VIM**.

Second, even if one granted Vaidya's claim about his version of **NOP**, one might avoid the problem by re-introducing a de-modalizing version of the principle which eliminates the modal auxiliary verb 'must':

NOP* An instance of **VIM** only yields a judgment about whether *P* is an essential property of *o* through property variation on *o* via the construction of a set of scenarios *S*₁...*S*_{*n*}, if *it is the case* that *o* is preserved in the transition from each *S*_{*i*} to *S*_{*k*}.

Since **NOP*** is a non-modal principle, assuming that the conditional involved is the material conditional, it cannot be considered an anti-luck condition, neither in the sense of a safety, nor in that of a sensitivity condition.³⁷ Still, it can block the circularity problem arising from **NOP**-violating scenarios. Just like **NOP**, **NOP*** makes the preservation of *o* throughout all scenarios involved in an instance of **VIM** a necessary condition for its success. The new principle could therefore still do its job and it could furthermore consistently be added as a success condition on Vaidya's version of **VIM**.

Re-introducing **NOP** or **NOP*** indeed saves Vaidya's proposal from the problem raised in the previous section. But does this addition result in a workable variant of Vaidya's proposal? There are three important reasons to doubt this.

³⁷ Anti-luck conditions in epistemology are usually spelled out in terms of a range of *closest* possible worlds and are therefore essentially modal: *Safety conditions* require that for a subject to know that *p*, *p* has to be true in the closest worlds in which it is believed (see e.g. Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 137–49, Pritchard, *Epistemic Luck*, ch. 6, Timothy Williamson, *Knowledge and its Limits* (Oxford: Oxford University Press, 2000), chs. 4–5, *sensitivity conditions* in contrast require that *p* is not believed by the subject in the closest *non-p* worlds (see e.g. Robert Nozick, *Philosophical Explanations* (Cambridge: The Belknap Press of Harvard University Press, 1981): ch. 3.

5.1 VIM's Quality as a Method for Acquiring Objectual Understanding of Essence

It is clear that Husserl considers **VIM** to be a method. To be more specific, a method "for the acquisition of pure concepts or concepts of essences."³⁸ Surprisingly, Vaidya never refer to **VIM** as a method in his "Understanding and Essence." Yet, he never explicitly denies that it is one either. Since it is very hard to see what else **VIM** could be, I will follow Husserl in assuming that **VIM** is a method.

The question which I will discuss in this subsection is whether **VIM** is a good method, a method which a subject seeking to acquire objectual understanding of essence could rationally adopt to pursue this goal. There are many different criteria for the overall quality of an epistemological method. I will here focus on two important quality criteria: First, the method's rate of success and second, what I will call its *transparency*. I call a method transparent, if it affords the subject who applies it feedback on whether it can terminate with success during different stages of its application. The main point of this subsection will be that **VIM** fares badly with respect to both of these criteria. My argument for this point will rely on a systematic connection between the two criteria.

Let me start with **VIM**'s transparency. A point made in the previous section was that a subject going through an instance of **VIM** cannot recognize a **NOP**-violating scenario as such. To do so, it would have to be able to recognize this scenario as resulting from the subtraction of an essential property of the object, because this is the only plausible way to guarantee that an application of **VIM** involving such a scenario could succeed. This, the subject cannot do, because this would require it to already have the sort of robust epistemic access to the essence of the relevant object which it is seeking to acquire by applying **VIM**. This point can be generalized to **NOP**-conform scenarios: In order to recognize a scenario as such, the subject would have to recognize that the property which it subtracted to create the scenario was *not* an essential property of the relevant object. This would also require the subject to already have prior epistemic access to the relevant object's essence which, by hypothesis, it does not have. So a subject who relies on **VIM** to acquire objectual understanding of essence can, on pain of epistemic circularity, neither recognize the scenarios which it creates while going through this process as **NOP**-violating, nor as **NOP**-conform.

NOP spells out a success-condition for **VIM**, a condition which has to be met for an instance of **VIM** to succeed. Since a subject cannot recognize of any scenario whether it conforms or fails to conform to this success condition, a subject

³⁸ Husserl, *Experience and Judgment*, 340.

applying the method cannot, at any stage of the process, tell whether its application of **VIM** meets this success-condition.

This generalizes an important point about the nature of Vaidya's proposal, which he explicitly mentions: Even if a subject manages to acquire objectual understanding of an object's essence through **VIM**, this epistemic state does not involve the subject's recognition that it has in fact attained this state. This follows directly from the account of objectual understanding which Vaidya accepts.³⁹ The point just made shows that this lack of recognition not only affects a subject after it has successfully applied **VIM**, but rather it completely pervades any instance of **VIM** at all of its stages, from the very first scenario on. We can conclude that **VIM** is lacking in transparency and fares badly with respect to this first quality-criterion.

Let me now focus on **VIM**'s rate of success. Concerning this factor, I have a speculative argument to offer. Consider the class of all instances of **VIM**. Can we say anything about how many of these instances may succeed? To be sure, we cannot give a precise answer since we can neither determine the total number, nor the number of the successful instances of **VIM**. There are however two things we know: First, only **NOP**/**NOP***-conform instances can succeed, and second, **VIM** is intransparent in that a subject going through an instance of the process cannot, at any stage, tell whether it conforms to **NOP**. These two point taken together with a third strongly suggest that **VIM** does not have a good rate of success. The third point is that following Husserl, **VIM** is designed to have "a structure of arbitrariness,"⁴⁰ which means that the process of producing scenarios in an instance of **VIM**, or to use Husserl's terminology, variants in an application of the method of essential seeing, is such that "it is a matter of indifference what might still be joined to it, a matter of indifference what, in addition, I might be given to apprehend in the consciousness that 'I could continue in this way.'"⁴¹ The underlying idea is of course that an object's essence will stay the same throughout all possible scenarios, no matter how arbitrary the changes are which the subject makes to produce them.

If we put this third point and the first two together, we get that **VIM** is a process which is very likely to fail: The process is designed to allow the subject to create scenarios by arbitrarily varying the properties of the object on which the scenarios are centred and it can only successfully terminate if these scenarios all conform to **NOP**. However, the subject cannot recognize whether the scenarios it

³⁹ See Vaidya, "Understanding and Essence," 827–828.

⁴⁰ Husserl, *Experience and Judgment*, 342.

⁴¹ *Ibid.*

creates conform to this success condition or whether they don't. The success of an application of VIM therefore appears to be mostly a matter of luck, since the subject has to create scenarios without any means to verify that they do not undermine the successful completion of the process. This does not show that **VIM** cannot succeed, but it strongly suggests that it is, taking into account two plausible quality criteria, not a good method for acquiring objectual understanding of essence.

5.2 The Contemporary View of Essence and the Apriority of VIM

Vaidya aims to develop a general epistemology of essence which fits the contemporary view of essence of philosophers such as Fine and Lowe.⁴² Following Kripke,⁴³ this view characteristically admits cases in which the essence of an entity is not discoverable by a priori means: The microstructure of a substance might for example be said to be essential to it, even though it is not discoverable by a priori means. More generally, the idea is that the two distinctions of it being imaginable or not that an object has a certain property and of it being essential or accidental to that object whether it has that property cut across each other. **VIM** faces a fundamental problem in accounting for knowledge of/understanding of an entity's essence if we assume this contemporary view of essence.

When describing the method of essential seeing, Husserl is clear that he assumes there to be a close connection between the thinkable and imaginable and the essential: "The essence proves to be that without which an object of a particular kind *cannot be thought*, i.e., without which the object *cannot be intuitively imagined as such*."⁴⁴ According to Husserl, "[w]e can direct our regard toward it[the essence] as toward the necessarily invariable, which prescribes limits to all variation practiced in the mode of the 'arbitrary.'"⁴⁵ With the contemporary notion of essence in mind, this quotation could be taken to give us a possible answer to our problem since one might understand it to say that while engaging in the imaginative activity required by **VIM**, the subject's imagination is limited by the relevant essence.⁴⁶ This limitation would then be what prevents the subject

⁴² See Vaidya, "Understanding and Essence," 819.

⁴³ See Kripke, *Naming and Necessity*.

⁴⁴ Husserl, *Experience and Judgment*, 341. My emphasis.

⁴⁵ *Ibid.*

⁴⁶ Note that this contradicts the interpretation of Husserl developed in De Santis, "Phenomenological Kaleidoscope," 31–33. Since my concern here is not with Husserl's own view of the method, but rather with a view which adapts it to the notion of essence of contemporary metaphysicians, I will only remark that this interpretation would not help in solving the

from producing a variation in which the relevant object lacks one of its essential properties. One might hence say that there is a pre-established harmony between what we can imagine about an object of a certain kind and its corresponding kind-essence. If we import this idea into the contemporary context, it amounts to what one may call *essential rationalism*.⁴⁷ A philosopher who accepts it assumes that the imagination-based method of essential seeing gives us a priori access to essence.

What **VIM** could hence offer a contemporary epistemologist of essence is a systematic method for probing which of an object's properties we can subtract from it in our imagination and which we cannot, *presupposing the essential rationalist view that what we can imagine about an object coincides with the possible states of the object left open by its essence*. This presupposition however directly conflicts with the contemporary view that imaginability and essentiality come apart. The question of how to bridge this gap, or the corresponding gap between conceivability and metaphysical possibility, is of course one of the core questions in contemporary epistemology of modality.⁴⁸ An answer to this question would be vital to Vaidya's proposal and more generally to any epistemology of essence based on the Husserlian method of essential seeing. It is not surprising that Husserl himself⁴⁹ provides no answer to this contemporary question, but we do not find it addressed in Vaidya's paper either. What is clear however is that from a contemporary perspective, the view about the relation between the imaginable and the essential which one can extract from Husserl falls short of providing a satisfying answer.

5.3 Conflation of Essence and *de re* Modality

A further problem with the modified version of Vaidya's proposal concerns the relation between the notions of essence and of metaphysical necessity, two notions which are clearly distinguished by contemporary essentialists.⁵⁰

Consider an instance of **VIM**. In it, the subject starts out with a particular experience or intuition of an object which is then subsequently modified by

problem at hand since it takes the subject who applies the method itself, not the relevant essence, to be the source of this limitation.

⁴⁷ Compare modal rationalism as e.g. characterized in Chalmers, "Does Conceivability Entail Possibility?" 172–173.

⁴⁸ See e.g. Chalmers, "Does Conceivability Entail Possibility?" and Peter Kung, "Imagining as a Guide to Possibility," *Philosophy and Phenomenological Research* 81, 3 (2010): 620–663.

⁴⁹ Who holds that "essential truths are called *a priori*; this means, *by reason of their validity, preceding all factuality*, all determinations arising from experience." (Husserl, *Experience and Judgment*, 352–353).

⁵⁰ I owe the main idea for this objection to Claudio Calosi and Fabrice Correia.

subtracting properties of the object to produce the plurality of variations which is then considered as a whole by the subject in order to see which properties remain constant throughout all of them. Both the subject's starting point and the variations produced can be represented, without giving up on anything which is essential to the intended workings of **VIM**, in terms of either sets of propositions or facts involving the relevant object. Since possible worlds correspond to maximally consistent sets of propositions or facts, this means that we can think of them as corresponding to parts of possible worlds. This should immediately give us pause.

Contemporary essentialists hold that essence is not definable in terms of necessity. The notion of metaphysical necessity (*de re*) can adequately be captured using possible worlds, but the notion of essence cannot be: Possible worlds are an adequate tool for capturing the metaphysically necessary properties of an object, but they cannot be used to capture the object's essence, since there are some properties which some objects have with metaphysical necessity, even though they do not belong to these object's essences: Socrates is necessarily distinct from the Eiffel Tower, but it is not essential to him that he is. This is of course the core insight of Fine's influential objections to the modal definition of essence.⁵¹

Since **VIM** can equivalently be restated as a method which relies only on possible worlds, this crucial difference means that there is in fact no good reason to think that the method can allow us to acquire (knowledge or) understanding of the essential, rather than of the necessary properties of an object. To put the point differently, from a contemporary perspective, essence is a hyperintensional notion, but Vaidya's variant of the Husserlian method is only intensional in nature. What we are looking for is a foundational method for the epistemology of essence, but it turns out that we have no good reason to think that **VIM** gives us such a method, as opposed to a method for acquiring (knowledge or) understanding of metaphysical necessity (*de re*).

6. Whither Essential Seeing and Objectual Understanding of Essence?

Vaidya's proposal combines two components which have been largely neglected in the contemporary discussions about the epistemologies of essence and of modality. The first is the Husserlian method of essential seeing, the second the idea of refocusing the debate on understanding instead of knowledge. The focus so far was on the combination of the two, but in this last section, I want to briefly comment on them separately in light of the previous discussion, starting with the Husserlian method/**VIM**.

⁵¹ See Fine, "Essence and Modality."

It is important to note that the problems for VIM which were raised in the previous section arise no matter whether the method aims at understanding or at knowledge. Since I do not see any salient alternatives to these two proposals, I believe that **VIM** cannot play the central role in the epistemology of essence which Vaidya allocates to it. This role requires a method which permits a subject to acquire arbitrary bits of objectual understanding (or knowledge) of essence without having any previous epistemic access to essence, something which **VIM** does not deliver.

That said, there still are two roles which it might be able to play. First, **VIM** appears to give us a viable method for determining, within certain limits, the essential properties of objects which belong to domains which are wholly accessible via our imagination and more generally via a priori methods. This may be the case for some geometrical objects, as Tieszen argues.⁵²

Second, **VIM** may give us a general systematic method for making tacit knowledge of essence explicit.⁵³ It is plausible that freely varying the properties of an object may help us get clear on which of an object's properties we assume to essentially belong to it. Assuming that we can acquire knowledge or understanding of essence in another way, **VIM** may thus help us realize which of our knowledge about an object concerns its essence. Another way in which the method may help us in this manner is as a device for inductively generalizing from knowledge or understanding of individual essences. A subject might for example apply **VIM** to different objects which share a certain essential property, noticing in each case that another property of these objects can likewise not coherently be subtracted when forming a new scenario. Based on this realization, the subject may inductively (and therefore defeasibly) infer that the two relevant properties are essentially connected with each other. This in turn may allow the subject to form the hypothesis that the second property is also essential to objects which have the first.

Can Vaidya's proposal to refocus the epistemology of essence on objectual understanding help solve problems faced by a knowledge-based epistemology? As I have stressed in the previous section, to have objectual understanding of an object's essence, a subject still needs to have true beliefs about its essence. In this respect, any epistemology of essence which builds on the former instead of the latter notion will still have to answer the question of how we can reliably acquire true

⁵² See Richard Tieszen, "Free Variation and the Intuition of Geometric Essences: Some Reflections on Phenomenology and Modern Geometry," *Philosophy and Phenomenological Research* 70, 1 (2005): 153–173.

⁵³ Thanks to Olivier Massin for this suggestion. An interpretation of Husserl's own method which points in this direction is proposed in Kasmiers, "Defense of Husserl's Method."

beliefs about essence. So it seems that the move to objectual understanding does not really help us address the most pressing fundamental question in the epistemology of essence. This problem is magnified by the fact that objectual understanding, as Kvanvig and Vaidya understand it, is by design unsuitable to give us a reliable method, since it is compatible with epistemic luck. Perhaps other conceptions of objectual understanding could fare better in this respect, but this is a topic for another day.⁵⁴

⁵⁴ I would like to thank participants of sessions of the research colloquium at the Institute de philosophie of the University of Neuchâtel and of the eidos seminar at the University of Geneva. Special thanks to Claudio Calosi, Donnchadh O'Connail, Fabrice Correia, Arturs Logins, Olivier Massin, Kevin Mulligan, Maria Scarpati, Peter Simons, Jonas Waechter and three anonymous reviewers. The first steps towards writing this paper were made several years ago in a section of my phd thesis for which I received funding from the European Community's Seventh Framework Programme FP7/2007-2013 under grant agreement no. FP7-238128. Further work on the paper was made possible by the SNSF-funded project 'Identity in Cognitive Science, Quantum Mechanics, and Metaphysics,' grant number 185435, Università della Svizzera italiana, PI: Kevin Mulligan.

EMPATHY AS A TOOL FOR LEARNING ABOUT EVALUATIVE FEATURES OF OBJECTS

Diana SOFRONIEVA

ABSTRACT: It is generally agreed that empathy can give us knowledge about others. However, the potential use of empathy as a tool to learn about features of objects in the world more generally, as opposed to learning only about others' internal states, has not been discussed in the literature. In this paper I make the claim that empathy can help us learn about evaluative features of objects in the world. I further defend this claim by comparing empathy to testimony. Then I present and respond to two possible objections to this analogy.

KEYWORDS: empathy, experiential imagination, testimony, emotional evaluation, epistemic practice, knowledge

1. Introduction

In this paper I will discuss one of the epistemic functions of imagining what it is like to be another person. I will call this imagining 'empathic.' Even though the way the term 'empathy' is used varies greatly,¹ this activity of imagining is in line with many accounts of empathy which see it primarily as an activity of the imagination.² What is important for the purpose of this paper is that I assume that empathy involves among other things imagining *what it is like* to be the other person, i.e. imaginings with a phenomenal content to them, that involve the realisation that to be the other feels like *this*. This is an example of what Dokic and Arcangeli refer to as "experiential" imagination: "we shall spell out the notion of experiential imagination as the imaginative capacity to re-create experiential perspectives."³ Experiential imagination is similar to Goldman's "enactment"

¹ Karsten Stueber, "Empathy," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Spring 2018 (Metaphysics Research Lab, Stanford University, 2018), <https://plato.stanford.edu/archives/spr2018/entries/empathy/>.

² There are other accounts, e.g. those that see empathy as an affective response not necessarily mediated by imagination, or a kind of approval and understanding of another person. However, the view of empathy as imagining is wide-spread.

³ Jérôme Dokic and Margherita Arcangeli, *The Heterogeneity of Experiential Imagination* (Open

imagination (or E-imagination) which is “a matter of creating or trying to create in one’s own mind a selected mental state, or at least a rough facsimile of such a state, through the faculty of the imagination.”^{4,5}

There has been no shortage of discussion in the literature about the epistemic function of empathic imagining with the goal of understanding another person. These discussions have been about the extent to which empathy can serve as a way to obtain knowledge of other people’s states. The question that has been discussed is to what extent we can rely on empathy to obtain knowledge about what the other person’s state is. Different accounts vary in their levels of optimism about the prospect of obtaining reliable information from empathy in this sense. But safe from the most pessimistic accounts it is agreed that via empathy we can learn something about the state of another person.⁶

But another topic seems to have received little attention in philosophy, so much so that the present author can find no reference to it in the literature. It is about the epistemic function of empathic imagining with the goal of obtaining knowledge about things in the world, beyond the state of the person we are empathising with.

In this paper I will argue that empathy allows us to obtain knowledge about objects in the world. It does so, because imagining another’s state involves imagining their evaluations of objects in the world and thus allows us access to these evaluations. To strengthen the claim that this practice is epistemically valuable, I will draw an analogy between empathy and testimony, which is an epistemically robust practice. Then I will raise and answer to two objections to the analogy. One is that testimony is not effective when it comes to evaluative

MIND. Frankfurt am Main: MIND Group, 2014), <https://doi.org/10.15502/9783958570085>.

⁴ Alvin I. Goldman, *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading* (Oxford University Press, 2006), 42, <http://www.oxfordscholarship.com/view/10.1093/0195138929.001.0001/acprof-9780195138924>.

⁵ Kind speaks of a similar kind of imagining, referring to it “recreative” imagining. Amy Kind, “Desire-Like Imagination,” *CMC Faculty Publications and Research*, 1 January 2016, http://scholarship.claremont.edu/cmc_fac_pub/533; see also Gregory Currie and Ian Ravenscroft, *Recreative Minds: Imagination in Philosophy and Psychology* (Oxford University Press, 2002).

⁶ Of course, this something depends on many factors, such as how much prior information we have about the person and their situation. If I have known you all my life I am likely to be better able to imagine what you are going through than if I just met you. Another factor is that people’s ability to empathise well varies greatly: we all know people who seem to incapable of getting even the most obvious distress, as well as others who can read us so well to the extent that we feel uncomfortably exposed in their presence.

properties of objects. The other is that testimony without a speech-act is at best a very weak form of testimony.

2. Empathy – Testimony Without the Middle Man

Empathy gives us information about the state of another person. Now I will make the case that it therefore also gives us information about features of objects⁷ in the world. The reason for that is that one's experiences are directly tied up with one's environment, and hence one's (emotional) state is usually closely connected to features of the external world. For the most part our emotions are directed at the external world. Emotions have intentionality; they are about objects, and represent the world as being in a certain way.⁸ They often reflect some kind of evaluation, or appraisal, one has of a certain object. In the most rough form possible: if I fear swimming in the river then I believe that (or judge that) swimming in the river is dangerous; if I trust Alex, then I believe that (or judge that) Alex is the kind of person that merits trust, and so on. Virtually all accounts of emotion agree that emotions represent the world as being a certain way.⁹

Emotional evaluations carry information about evaluative properties of the object the emotion is directed to. They are evaluations with a phenomenal feel to them.¹⁰ Correspondingly, evaluative properties of objects are properties "whose recognition merits a certain sort of response."¹¹ For example, dangerousness is an evaluative property; the response merited by something dangerous is fear "with all that this emotional experience involves, including thought, feeling, and action."¹² If swimming in the river is evaluated as dangerous, this carries the information that fear is the appropriate response to swimming in the river.

⁷ I use 'objects' in the broadest possible sense: what counts as an object would involve physical objects, but also ideas, people, situations, and other things.

⁸ Andrea Scarantino and Ronald de Sousa, "Emotion," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2018 (Metaphysics Research Lab, Stanford University, 2018), <https://plato.stanford.edu/archives/win2018/entries/emotion/>.

⁹ This is in contrast to the now largely rejected "dumb view" on emotions, according to which emotions are just feelings. See, e.g. Alison M. Jaggar, "Love and Knowledge: Emotion in Feminist Epistemology," *Inquiry* 32, 2 (January 1989): 151–76, <https://doi.org/10.1080/00201748908602185>.

¹⁰ I am neutral on the question of whether, apart from emotional evaluations, there are other kinds of evaluations of objects which have a distinctive phenomenological feel to them, or whether all such evaluations could be considered emotional evaluations.

¹¹ Peter Goldie, *The Emotions: A Philosophical Exploration* (Oxford: Clarendon Press, 2002), 30.

¹² Goldie, 30.

Different accounts of emotion offer different answers to how the feel of an emotion relates to the emotional evaluation. For example, according to a view advocated by Prinz emotions are embodied appraisals – they represent objects in the environment, and the phenomenal feel of an emotion is what makes us aware of this representation.¹³ Emotions are not compounds of judgments and embodied appraisals, but rather “embodied appraisals that have been recalibrated by judgments to represent somewhat different relations to the environment.”¹⁴ But on most accounts of emotion, emotional evaluations are in some way tied to an emotional experience’s phenomenological feel, and the way an emotion ‘feels’ carries some information about the corresponding emotional evaluation.¹⁵

If empathy involves imagining how you feel about a certain thing, then empathy allows me access to your emotional evaluations of features in the environment you are in. From this I can learn something about this environment. For instance, if via empathising I figure out that you are afraid of the upcoming exam, then I have also learned that the upcoming exam is on material you are not well prepared for. In almost all cases where we learn something about others via empathy, the other side of the coin is that we have learned something about the environment these others are embedded in and have enriched our understanding of the world beyond the state of the particular person whose situation we had imagined.

Importantly, what I can learn via empathy here is limited to evaluative features of objects. There are a lot of things that I cannot learn via empathy, as for example, physical facts about objects like the fact that the chair is green or that water is H₂O. I cannot learn these things because they are not the sort of things that would normally evoke an evaluative response in you which has a distinctive phenomenal feel. Therefore, experientially imagining your experience is unlikely to give me information about such things. I can only learn similar facts from you if you tell them to me explicitly. What I can learn via empathising with you concerns features of objects that would provoke an evaluative response in you with certain distinctive phenomenology, as for example, emotional response: I can learn that a certain work of art or a certain person’s deed is admirable, that swimming in the river is fearsome, that a certain meal is disgusting, and so on.

¹³ Jesse J. Prinz, *Gut Reactions: A Perceptual Theory of Emotion*, Philosophy of Mind (Oxford, New York: Oxford University Press, 2006).

¹⁴ Prinz, 99.

¹⁵ An emotion’s phenomenological feel is an important aspect of emotional experience – this is so on virtually all accounts of emotion, excluding what Prinz calls “pure cognitive theories.” (Prinz, 10).

I am suggesting that the inference occurs in two steps. First, I obtain knowledge what a certain object, *P*, is for you. Then I infer what *P* might be for me. In the first step I experientially imagine your state which involves imagining a particular phenomenal *feel* of the emotion you are undergoing – imagining that you feel like *this*. In this way, I have access to your emotional evaluation of the object. Then, in the second step, I use your evaluation in order to infer the corresponding evaluative features of the object.

For example, if you are telling me about an episode of confrontation with one of your colleagues, when I empathise with you I will figure out that you felt like *this* about them. What I have already learned is that your colleague has the sort of features to which *you* would react in this way. From here I might be also able to take a second step and form a picture of what kind of features your colleague is likely to ‘objectively’ possess, or if one doesn’t want to commit to there being objective traits in this sense, I can infer what kind of features *I* might perceive your colleague as having.

In order for the second step in the inference to be successful I need to know enough about the extent to which you are likely to react to the relevant objects in a similar way as myself. Sometimes this might be a relatively ambitious task, for example when the emotional evaluation is directed towards a person. If I form the judgment that you feel like *this* towards Ben, unless I know you very well, I would not be able to infer reliably almost anything about Ben, because of the complex ways in which people’s characters react to one another. At other times, however, when the object in question is a physical object in the environment to which most people are likely to have similar evaluative responses, the possibility of inferring things reliably about it is not so far-fetched. For example, if you feel some kind of unease, about visiting a particular bar, perhaps I do not need to know you too well to be able to infer that I, too, am likely to feel uneasy about visiting that place, and to also consider it dodgy. Since I am not making a claim about how often we can rely on empathy to learn about objects in the world in this way, the fact that there are complex cases where the inference is not likely to be successful unless I know you well, is not a counterexample to what I am trying to show. All I am trying to show is that via empathy we make use of other epistemic agents’ understanding of the world, and we can appropriate it (or adjust it) to expand our own understanding.

Now a question that suggests itself is the extent to which other epistemic agents’ evaluations of objects in the world would be epistemically useful to us. I will defend the claim that it is useful by drawing a parallel between empathy and testimony.

Testimony is an ubiquitous source of knowledge and our dependence on it is far-reaching.¹⁶ It is a robust epistemic practice, and is one of our sources of knowledge together with perception, memory, and reasoning. We usually accept ordinary informative testimony – normally it is infeasible “for hearers to seriously check or confirm either the speaker’s reliability or sincerity within the normal constraints of testimonial transmission and exchange.”¹⁷ It is an open question under what conditions testimony is a justified source of knowledge. Non-reductionists hold that what is required is only the absence of undefeated defeaters whereas according to reductionists some actual positive reasons are necessary too in order to accept the testimony of speakers.¹⁸ However, it is universally accepted that we can, and that we do, attain knowledge from what others tell us.

The parallel between empathy and testimony is the following. In both cases I have not attained the knowledge of an object first-hand. In the case of testimony, I rely on your evaluation of X, which you share with me via a speech act. In the case of empathy, I rely on your evaluation of X, which I access via experientially imagining your state. In this sense empathy is like testimony without the middle man. In testimony I rely upon your assertion in order to access your knowledge about an object. In empathy you need not tell me anything – I rely upon my experientially imagining your state, in order to gain epistemic access to your emotional evaluation of the object. And if we accept that testimony is sufficiently often a reliable source of epistemic benefits, we should be justified to accept the same about empathy. If it is undeniable that our practice of testimony is epistemically robust, and if it is true that there is an analogy between empathy and testimony, then it would seem that arriving at epistemic benefits via empathy would be epistemically justified. The hope here is that whatever it is that justifies testimony as a source of knowledge, can justify why empathy too is a source of knowledge. In both cases your evaluative knowledge of certain objects in the environment is transmitted to me – it is just that it happens without your explicit assertion.

¹⁶ Lackey Jennifer, "Knowing from Testimony," *Philosophy Compass* 1, 5 (2006): 432–48, <https://doi.org/10.1111/j.1747-9991.2006.00035.x>; Jonathan Adler, "Epistemological Problems of Testimony," in *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta, Winter 2017 (Metaphysics Research Lab, Stanford University, 2017), <https://plato.stanford.edu/archives/win2017/entries/testimony-episprob/>.

¹⁷ Adler, "Epistemological Problems of Testimony."

¹⁸ Lackey Jennifer, "Knowing from Testimony."

3. Objections: Empathy Is Unlike Testimony

However, there are important differences between empathy and testimony and one might wonder whether the analogy between the two is justified. If it is not justified, one might wonder to what extent I could be justified to infer, from the fact that *you* evaluate an object X in a way E, that X is an appropriate object of such an evaluation. There are two reasons one might reject the analogy. I will consider these challenges now.

3.1 Objection 1 – Testimony About Evaluative Properties Is Problematic

The first reason the analogy between empathy and testimony might fail is that there is an important difference between the domain of things I can learn by empathy and the domain of things it is usually agreed I can learn via testimony. Via testimony I can attain knowledge of facts – that water is H₂O or that the museum is behind the corner. However, this is not the sort of thing I can learn via empathy. Via empathy I can only attain knowledge about evaluative features of objects. Hence empathy is at best like *testimony about evaluative properties* only. And one might wonder about the extent to which evaluative testimony – unlike testimony for physical properties of mid-sized objects, directions in a city, or train times – can be a reliable and justified epistemic practice. If testimony is not a reliable and justified source of epistemic goods in this sense, then it is not clear why empathy should be one. Perhaps evaluative aspects of objects are just not the sort of thing about which we can rely on attaining knowledge second-hand, be it via testimony or empathy.

There is an essential dissimilarity between beliefs and evaluations. Evaluations are more subjective and personal, whereas beliefs are more objective. Whereas I can take your word for statements such as “the museum is behind the corner” and “water is H₂O,” perhaps I cannot take your word for statements such as “the picture is beautiful” and “John is trustworthy” since these evaluations would have an important subjective element. They are the kind of things that it is possible that are true for you, but might not be true for me. Because of this difference between the two, it seems that evaluative knowledge is less straightforward to attain second-hand than knowledge of facts. Whereas I can simply rely on your testimony that the museum is behind the corner, it is less clear why I should accept your statement that the movie is imaginative. The appropriateness of an object evaluation is relative to the person who evaluates it. And one might worry that in some cases people’s emotions towards the same object can differ drastically in a way in which we would not normally expect their beliefs to differ.

In short, the worry is that from your evaluating an object in a certain way I cannot simply infer that I should evaluate it in the same way because of the subjective element that you bring with your evaluation. However, it seems that I can still attain knowledge about that object because I might be able to accommodate the subjective element via taking into account relevant differences between you and me, and infer what kind of evaluation I should be having of that object. Even if it is more complex than assessing the reliability of a speaker in testimony, it does seem possible that I might be able to infer from evaluations you are having about certain objects, what evaluations I should be having about these objects.

Taking another's evaluation to reflect an object's features might not be as straightforward as taking their beliefs concerning that object to reflect an object's features, but at least in some cases I can adjust for differences between you and me. If I know sufficiently many things about you, I would know for example where you and I differ, and what adjustments I need to make to your evaluations in order for me to be able to accept them as epistemic evidence of certain evaluative aspects of objects. I might have independent reasons to discredit your evaluation. Say, for example, that you are terrified by a spider in the kitchen. If you have a phobia of spiders and I know you, I would know that you have a phobia, and I would not take your emotional evaluation as direct evidence of the scariness of that spider. The opposite can also be true – I might have independent reasons to allocate especially large epistemic credibility to your evaluation. If I know that you are a zoologist without phobias, and if you are particularly alarmed by a spider in the kitchen, I would infer that the spider in the kitchen is perhaps a poisonous one and I should beware it.

Perhaps the least controversial case of me learning about the world via empathy is the case where I know you particularly well. Say you are my best friend, my sibling, my partner, or someone else who is really close. If I know you sufficiently well it seems that it will be very easy for me to infer from your emotional evaluations whether or not certain features about the objects of these evaluations hold. Say you and I are very similar in our appraisals of people – we find particular features in people morally repugnant, we hold particular values in high regard, and so on. If you react emotionally to a certain person I do not know in a particular way – say, with a kind of derision, or admiration – I will know that this person is more likely than not a person whom I myself would consider an appropriate object of derision, or admiration. I doubt it that someone would deny that in *these* specific cases, where we empathise with people we know very well, we can learn something about the world from them.

The question now is to what extent this learning can happen in other cases. One might object that in order to do the appropriate adjustments reliably I need to know you well. This might be true for more complex evaluations such as those concerning other people's characters, where it seems that I indeed need to know you well in order to interpret your evaluation in a sensible way. However, for more basic evaluations such as responses to physically dangerous objects, it seems that I can take on board your evaluation even if I do not know you well. But even if empathy works well as a source of epistemic goods about the world in the limited cases of people we know well this doesn't mean empathy isn't in this sense epistemically valuable, since we tend to interact a lot more with people who are closer to us and we know well. So those are two reasons to believe that attaining knowledge about evaluative features of objects in the world via empathy might be more pervasive than it initially seems. But even so my claim is a weak one – insofar as I have shown that learning about the world via empathy happens at least sometimes, I have shown what I was aiming to.

3.2 Objection 2 – Testimony Takes Its Epistemic Credentials From the Speech-act

There is another reason to doubt that the analogy between testimony and empathy holds. One can hold that that which makes testimony a justified source of knowledge is essentially linked to the speech act of assertion in testimony. If that is true, then there is a crucial disanalogy between testimony and empathy – the difference between the two is not just an incidental, but an essential property of what makes testimony the epistemically robust practice that it is. Therefore, one would not be able to infer, from the fact that testimony is a justified epistemic practice, that empathy could be one too.

What is sometimes called the “assurance view” of testimony is the view according to which testimonial knowledge is warranted because the speech-act itself is assurance that what the testifier says is true.¹⁹ There is some attraction to the view that it is the norms of the conversational practice, and something about assuming responsibility for what I have said to you, that makes testimony reliable:

To use Kant's example: If I start to pack my suitcase in front of you, but I have no plan to leave then I intentionally deceive you by giving you evidence that I plan to leave. But I do not invite you to notice or to understand what I am doing. By contrast, if I said to you either ‘I am leaving town’ (a lie) or ‘Do not worry if you do not find me here tomorrow’ (an intentionally misleading assertion), I do invite

¹⁹ Nickel, Philip et al., "Assurance Views of Testimony," in *Routledge Handbook of Social Epistemology* (Routledge Taylor & Francis Group, 2019), 96–102, <https://doi.org/10.4324/9781315717937>.

you to understand and believe me. Thereby, I assume responsibility for the truth and veracity of my assertion, though arguably less so for the implicature that 'I am leaving town.'²⁰

There is no equivalent to this responsibility-taking in the case of empathy. I have no responsibility to have certain emotional evaluations of objects, at least not in the way in which I might be assumed to have in the case of asserting true statements. I might feel whatever I want to feel and have whatever evaluations I have of certain objects, and I will not be sanctioned in any way for doing that. By contrast, assertion of something untrue is generally sanctioned and assertion has to abide by certain norms of conversational practice that imply not deceiving or providing unjustified information. Hence it might seem that unlike relying on others' assertions, sourcing knowledge from others' emotional evaluations, is not a reliable source of epistemic goods.

One line of reply might be to try to argue that there are some social norms of emotional evaluation. For example, we usually don't trust people who display a different emotion from the one they are actually experiencing. We like and value spontaneous people and value genuine emotional expressivity and people who are sincere with their emotions.²¹ So one might think there is some kind of equivalent of the norms of conversational practice in the emotional evaluation domain. We read each other's emotions all the time and to a certain extent we rely on others' evaluations being adequate, and this does make it likely that via empathy I would attain knowledge of objects.

Another line of response is to turn the objection on its head – to agree that there is a disanalogy between cases of learning via testimony and cases of learning via empathy, but to claim that it is *precisely* in the disanalogy between speaking and emotions, that makes learning via empathy a valuable source of knowledge. Let me elaborate. As a speaker I know that there are certain rules of conversational practice by which I better abide. I know that you, as a hearer, are on the look-out for what I say. This might make me carefully calculate what I say, and thus, by merely listening to what I say, you might not be able to get at what I *really believe*. Empathy, by contrast, offers us more of a window into what your actual evaluations of things are. You might say one thing but feel another, and if I do not empathise with you, I will not get what it is that you actually feel, which seems an important part of your evaluative judgment of things. This is based on the

²⁰ Adler, "Epistemological Problems of Testimony."

²¹ According to some, one of the reasons spontaneity is valued is because it is hard to fake, e.g. Edward Slingerland, *Trying Not to Try: The Art and Science of Spontaneity* (New York: Crown Publishing Group, 2014).

assumptions that (1) it is harder to fake an emotion than assert an untruth, since the former requires to put on an act, which is very involving, and most of us are not very well trained in controlling carefully our emotional expressions and our body language, among others; and (2) that I will not be so careful in orchestrating my emotions, or even think about doing it, because emotions are not part of a conversational practice in the same way as words are, so I would not be so likely to begin monitoring them in the first place. If these assumptions are any likely, in empathic transmission of information you have less opportunity to deceive me and empathy would allow me an insight into your honest evaluations of things, which I might not get at merely via testimony.

Now one might perhaps wonder to what extent it is valuable for me to know your *honest* evaluations of things, as opposed to the ones you would have shared with me merely by speaking? I think that most of us would intuitively think that it is obvious that we want to know what people really think, but perhaps one might object to this. There are cases where one does not know something very well. Perhaps you honestly believe what you tell me, it is just that you cannot bring yourself to feel about it in a certain way. Say you have become totally convinced that a certain art-work is a work of genius – it is just that you do not really feel it. Or that you know what ‘the right thing to do is’ – to be polite and accepting of someone’s view – it is just that you do not really feel you should do anything of the kind, perhaps the person in fact annoys you greatly. Now, will there be any value in my getting at these evaluations of yours? Perhaps it is undeniable that getting at them makes me know *you* better. But is getting at them valuable in terms of me learning something about the *objects* of your evaluation – of the art-work, and of what the right thing to do is in confrontation with that person? Since it seems possible that one’s emotional evaluations might lead someone astray, whereas one’s ‘purely’ reason-based evaluations are more often guaranteed to be on the right track, one might be seriously worried by these cases. However, it seems that even people who have this worry would be justified in holding that one is better for *not knowing* what another’s actual emotional response to something is.

A third way to reply to the initial objection is to deny the claim that testimony takes all its epistemic credentials from the speech act. Instead, one can hold that testimony is at best partly justified by the existence of a conversational practice and what makes testimony justified is that via it we get access to the speaker’s knowledge. What makes this a justified way to attain knowledge is something about people being in general good enough epistemic agents. If this is true, it seems that one would need further justification to argue that they aren’t in general good enough evaluators. In further support of the view that the existence

of norms in conversation does not add much to the credibility of testimony, one can take the fact that in many ordinary cases of testimony, the speaker rarely thinks a few times before they produce an assertion. On the contrary, many of the assertions that we take to be good examples of testimony are often spontaneous. For example, most of us *would*, I think, take as a reliable piece of testimony a statement if it was one we overheard someone speaking to themselves. So there is no special step that the speaker takes in order to become intentionally a part of the social conversational practice. There is no reason, in many ordinary cases of testimony, to believe that something qualitatively different happens when the speaker is speaking to you, rather than were they merely to assert something to themselves aloud, or were they to write their words down in a notebook they do not intend for anyone to read. This is not an altogether strange view of what testimony is. For example, Sosa takes that all testimony requires “a statement of someone's thoughts or beliefs, which they might direct to the world at large and to no one in particular.”²² Hence it seems that it would be awkward to place *all epistemic justification* in testimony on the speech-act. In other words, if a conversational practice did not exist, but we were able to directly tap into people's beliefs about certain matters, that would be a reliable source of epistemic justification. Therefore, it seems hard to deny that testimony works not *only* because people abide by certain conversational practices alone but also because whatever it is that the speaker has ‘in stock’ in his mind, ready for assertion, will be in general reliable. Further, it is hard to see how even the most rigid conversational practice would be able to produce epistemic goods, if the latter were not the case. And if this claim is true it is hard to see why the same would not hold for the case of emotional evaluations. If we can accept that people are good enough sources of beliefs and epistemic knowledge, why deny that they are good enough sources of evaluations? In other words, it seems difficult to doubt that, in general, most people would have roughly appropriate emotional evaluations of certain objects a lot of the time. Hence accessing these evaluations and using them to inform our own evaluations of these objects seems an important epistemic function that, I hope to have shown, empathy can perform.

4. Conclusion

In this paper I argued that empathy is a valuable tool for understanding others, and for understanding objects in the world beyond the experiences of other people. I argued that empathising with another allows us to understand objects in the world

²² Ernest Sosa, *Knowledge in Perspective* (Cambridge: Cambridge University Press, 1991), 219.

Empathy as a Tool for Learning about Evaluative Features of Objects

via giving us access to their emotional evaluations of these objects; I compared empathy to testimony and argued that if we consider testimony as a source of epistemic goods, then we can consider empathy to be one too.

TRUTH TRACKING AND KNOWLEDGE FROM VIRTUAL REALITY

Billy WHEELER

ABSTRACT: Is it possible to gain knowledge about the real world based solely on experiences in virtual reality? According to one influential theory of knowledge, you cannot. Robert Nozick's truth-tracking theory requires that, in addition to a belief being true, it must also be sensitive to the truth. Yet beliefs formed in virtual reality are not sensitive: in the nearest possible world where P is false, you would have continued to believe that P. This is problematic because there is increasing awareness from philosophers and technologists that virtual reality is an important way in which we can arrive at beliefs and knowledge about the world. Here I argue that a suitably modified version of Nozick's sensitivity condition is able to account for knowledge from virtual reality.

KEYWORDS: virtual reality, truth tracking, sensitivity, reliabilism, externalism

1. Introduction

Suppose S comes to believe that P based solely on their experiences in virtual reality. Is it possible for S to know that P? According to one influential theory of knowledge, they cannot. Robert Nozick's famous truth-tracking analysis requires that, in addition to a belief being true, it must also be sensitive to the truth:¹

Sensitivity: If it had not been that case that P, then S would not have believed that P.

Now it looks like any belief formed solely on the basis of virtual reality will fail sensitivity. In the nearest possible world where P is false, S would have had the same experiences, and therefore S would have continued to have believed that P. Whilst S's belief that P might be sensitive to the *virtual* world, it will fail to be

¹ In addition to the truth of P, S believing that P, and the sensitivity principle, Nozick's full analysis requires a fourth condition called 'adherence:' if it had been the case that P, then S would have believed that P. Given that the main problem for knowledge from virtual reality stems from the sensitivity principle, this will be the main focus of the paper.

sensitive to the *real* world. Nozick himself recognized as much when he discussed his account in connection with the famous brain-in-a-vat thought experiment:

There remains, for example, the case of the person in the tank who is brought to believe, by direct electrical and chemical stimulation of his brain, that he is in the tank and is being brought to believe things in this way. The person in the tank does not know he is there, because his belief is not sensitive to the truth... The operators of the tank could have produced any belief, including the false belief that he wasn't in the tank; if they had, he would have believed that.²

Nozick does not consider this a problem because he does not believe that a person in the brain-in-a-vat scenario is capable of knowledge. Even if the scientist decided to be honest and reveal truths to the envatted person, they would still not have knowledge—if the scientist had induced the same beliefs in a world where they were false, the envatted person would have continued to believe them.

However, Nozick's refusal to permit knowledge in 'virtual worlds' is beginning to look increasingly untenable. There are two main reasons for this. Firstly, philosophers have started to recognize the value of virtual reality as a source of belief and knowledge about the world.³ Writers such as Jon Cogburn and Mark Silcox,⁴ for example, make a comparison between virtual reality and other fictional media, such as novels, movies and computer games. Although fictional, these media can contain truth, and under the right conditions can provide knowledge about the world.⁵

Secondly, and perhaps more importantly, there is growing empirical evidence for the claim that knowledge can be attained from virtual reality. For at least five decades, virtual reality has been used in some form or another to train

² Robert, Nozick, *Philosophical Explanations* (Cambridge, MA: Belknap Press, 1981), 175.

³ See for example Jon Cogburn and Mark Silcox, "Against Brain-in-a-Vatism: On the Value of Virtual Reality," *Philosophy & Technology* 27, 3 (2014): 561-579; Eva Dadlez, "Virtual Reality and 'Knowing What It's Like': The Epistemic Upside of Experience Machines," in *Experience Machines: The Philosophy of Virtual Worlds*, ed. Mark Silcox (Lanham: Rowman & Littlefield International, 2017), 75-86; James McBain, "Epistemic Lives and Knowing in Virtual Worlds," in *Experience Machines: The Philosophy of Virtual Worlds*, 155-168.

⁴ Cogburn and Silcox, "Against Brain-in-a-Vatism," 561-579.

⁵ For the wider discussion of knowledge from fiction (which does not differentiate virtual reality from other forms of fictional media) see Axel Spree, "Fiction, Truth and Knowledge" in *From Logic to Art: Themes from Nelson Goodman*, eds. Gerhard Ernst, Jakob Steinbrenner and Oliver Scholz (Paris: Walter de Gruyter, 2009), 329-344; Kathleen Stock, "Learning from Fiction and Theories of Fictional Content," *Teorema: Revista Internacional de Filosofía* 35, 3 (2016): 69-85; AsbjørnSteglich-Petersen, "Fictional Persuasion and the Nature of Belief," in *Art and Belief*, eds. Ema Sullivan-Bissett, Helen Bradley, and Paul Noordhof (Oxford: Oxford University Press, 2017), 174-193.

individuals in the skills needed to fly planes, land spacecraft, and perform certain surgical operations. More recently, so-called 'educational VR' is being touted as a replacement to traditional teacher and textbook-led instruction. The aim is to create virtual worlds that represent objects and events that are difficult to explore experientially in the real world. There is significant evidence that individuals can come to a better understanding and gain new knowledge about objects and their behaviour in the real world as a result.⁶

If knowledge is attainable from experience in virtual reality, then it would show that Nozick's sensitivity condition is not necessary for knowledge. Virtual reality would join a list other belief-forming methods, such as induction, introspection and testimony that—whilst commonly believed to provide knowledge—nonetheless fail to meet sensitivity.⁷ However, just as it has been argued that the appearance of insensitivity in these methods only emerges when the sensitivity principle has been improperly applied, so I will argue much the same is true for virtual reality.⁸ Although Nozick's original sensitivity condition is too strong to account for knowledge from virtual reality, a suitably modified version of it—one that takes into consideration stages of belief formation within virtual worlds—can account for the correct cases in which knowledge is attained.

⁶ For examples see the discussion in section 2.

⁷ The case from induction has been discussed by Jonathan Vogel, "Tracking, Closure, and Inductive Knowledge," in *The Possibility of Knowledge: Nozick and His Critics*, ed. Steven Luper-Foy (London: Rowman & Littlefield, 1987), 197–215; Ernest Sosa, "How to Defeat Opposition to Moore," *Philosophical Perspectives* 13 (1999): 137–49; Duncan Pritchard, "In Defence of Modest Anti-Luck Epistemology," in *The Sensitivity Principle in Epistemology*, eds. Kelly Becker and Tim Black (Cambridge: Cambridge University Press, 2012), 173–192. The case from introspection has been given by Jonathan Vogel, "Reliabilism Leveled," *Journal of Philosophy* 97, 11 (2000): 602–623; Ernest Sosa, "Rational Intuition: Bealer on its Nature and Epistemic Status," *Philosophical Studies* 81, 3–2 (1996): 151–162; Ernest Sosa, "Tracking, Competence, and Knowledge," in *The Oxford Handbook of Epistemology*, ed. Paul Moser (Oxford: Oxford University Press, 2002), 264–287. For a discussion of testimony see Stanford Goldberg, "Sensitivity from Others," in *The Sensitivity Principle in Epistemology*, eds. Kelly Becker and Tim Black (Cambridge: Cambridge University Press, 2012), 43–65; Tristan Haze, "Two New Counterexamples to the Truth-Tracking Theory of Knowledge," *Logos & Episteme* 6, 3 (2015): 309–311.

⁸ Alternative versions of the sensitivity principle that have been given to solve these and other problems can be found in Joseph Salerno, "Truth Tracking and the Problem of Reflective Knowledge," in *Knowledge and Skepticism*, eds. Joseph Campbell, Michael O'Rourke, and Harry Silverstein (Cambridge, MA: MIT Press, 2010), 72–81; Goldberg, "Sensitivity," 43–65; Fred Adams, John Barker, and Murray Clark, "Knowledge as Fact-Tracking True Belief," *Manuscripta* 40, 4 (2017): 1–30; Kevin Wallbridge, "Sensitivity, Induction, and Miracles," *Australasian Journal of Philosophy* 96, 1 (2018): 118–126.

In the next section I start by defining more clearly what I mean by virtual reality and the kinds of experiences that I am concerned with as a method of belief formation. James McBain⁹ has recently proposed an account of knowledge in virtual reality as a response to what he considers are the flaws in Nozick's own theory. He uses Dretske's information-theoretic account in order to do so, but as I shall show, the central idea can be captured using a variation of sensitivity, what I call 'virtual sensitivity.' In section 4 I outline two objections to the McBain-inspired virtual sensitivity principle that show it is neither necessary nor sufficient for knowledge from virtual reality. Finally in section 5 I outline a new sensitivity principle, 'virtual sensitivity+', that I argue overcomes the problems with McBain's account and can explain the cases in which knowledge from virtual reality is attainable.

2. Virtual Reality as a Belief-Formation Method

In the original presentation of his tracking theory Nozick realized that the sensitivity principle needed to be relativized to the method through which the belief was arrived at. He illustrates this with his well-known 'grandmother case':

A grandmother sees her grandson is well when he comes to visit; but if he were sick or dead, others would tell her he was well to spare her upset. Yet this does not mean she doesn't know he is well (or at least ambulatory) when she sees him. Clearly, we must restate our conditions to take explicit account of the ways and methods of arriving at belief.¹⁰

This example fails sensitivity even though intuitively the grandmother has knowledge. In the nearest possible world where *P* is false (where *P* = 'her grandson is well'), the grandmother would continue to believe that *P*. In this possible world the grandmother would have believed it using a different method. Instead of using perception she would have based her belief on testimony. It seems reasonable then that when judging whether or not sensitivity has been satisfied, we must keep constant the method being used to arrive at a belief.

Although I won't argue for it here, I believe there are good grounds for treating virtual reality as a distinct method by which we can come to arrive at beliefs—one that depends causally on other cognitive faculties (much like testimony does) but sits somewhere between testimony and instrument-based belief.¹¹ For better or worse, Nozick himself thinks we can individuate methods

⁹ McBain, "Epistemic Lives", 155–168.

¹⁰ Nozick, *Philosophical Investigations*, 179.

¹¹ It is feasible, for example, that a virtual world could be designed and constructed in a completely automated fashion. Imagine a space probe that scans and maps some distant planet

simply based on our experience of them and how distinct they ‘feel’ to us.¹² For most it will not be obvious that virtual reality is a distinct method since few have (yet) had the chance to experience it. So what I aim to do here is to give some examples of the way virtual reality is currently being used, especially in training and education contexts, and provide a general outline of the main steps and cognitive processes used in arriving at beliefs through it.

It will be useful to begin with a definition of virtual reality. Although not uncontentious, I will follow the most widely held definition given by Howard Rheingold¹³ and Michael Heim¹⁴ according to which virtual reality is a computer-generated sensory experience that is both *immersive* and *interactive*. Immersion is a difficult idea to define precisely but for present purposes we can think of it as the subjective feeling of presence inside the virtual world produced by a computer. These feelings are generated by experiencing a sensory interface that can involve a range of technologies including: head-mounted displays, virtual reality rooms, surround sound headphones, and haptic equipment (such as suits and gloves) that provide feelings of force, motion, and even temperature to the user. What also separates virtual reality from other types of immersive media (like movies) is its interactivity. By making decisions via an input device, the user can change the outcome of the experiences that are being fed to them by the computer.¹⁵

In what ways can virtual reality be used to arrive at new beliefs about the world? Training simulators that use virtual reality have been used for decades in industries such as the military, healthcare, and aerospace.¹⁶ In most of these cases

and sends the data back to earth that is then automatically rendered into an immersive 3D virtual world for scientists to explore. This seems distinct from testimony and yet in an important sense depends on the human design of the hardware to reliably gather data and portray the distant planet. Whether or not virtual reality is a distinct method of belief-formation from testimony or instrument-based belief will not matter much for my argument. What matters is that beliefs about the real world can be generated from experiences in this way.

¹² Nozick, *Philosophical Investigations*, 184.

¹³ Howard Rheingold, *Virtual Reality: The Revolutionary Technology of Computer-Generated Artificial Worlds - and How It Promises to Transform Society* (New York: Simon & Schuster, 1992).

¹⁴ Michael Heim, *The Metaphysics of Virtual Reality* (Oxford: Oxford University Press, 1993).

¹⁵ These conditions provide a broad definition of virtual reality that will include not only state-of-the-art forms of virtual reality that utilize head-mounted displays, but also more familiar interactive media such as video games and training simulators. More narrow definitions are possible, but in these cases one needs to specify either precisely the hardware involved or the level of immersion produced. I will follow others who write on this issue by sticking to the broad definition, even if it includes experiences we do not normally call ‘virtual reality.’

¹⁶ Derek Stanovsky, “Virtual Reality,” in *The Blackwell Guide to the Philosophy of Computing*

virtual reality is used to train individuals in the skills needed to operate complex equipment, such as an aircraft or space probe. According to education theorists, what makes virtual reality so good at this is that it provides ‘situated learning’ opportunities that are difficult to have in the real world due to cost and safety concerns.¹⁷ The kind of knowledge that is gained from training simulators is skills-based knowledge or ‘knowledge-how.’ But more recent virtual reality programs have been created that aim to provide factual knowledge or ‘knowledge-that.’ The developers of these programs aim to replicate the advantages of situated learning that have been found in training simulators by applying virtual reality to more factual learning outcomes. Examples of programs that have already been developed include:

- River City (medicine and epidemiology)¹⁸
- Supercharged! (electrostatic forces)¹⁹
- Virtual Cell (cell biology)²⁰
- Immune Attack! (immunology)²¹
- Whyville (basic scientific concepts)²²
- Quest Atlantis (history)²³
- EcoMUVE (ecosystems)²⁴

and Information, ed. Luciano Floridi (Oxford: Blackwell, 2004): 167–177.

¹⁷ Christian Schott and Stephen Marshall, “Virtual Reality and Situated Experiential Education: A Conceptualization and Exploratory Trial,” *Journal of Computer Assisted Learning* (2018): 1–10.

¹⁸ Chris Dede, “Immersive Interfaces for Engagement and Learning,” *Science* 323, 66 (2009): 66–69.

¹⁹ Janice Anderson and Mike Barnett, “Learning Physics with Digital Game Simulations in Middle School Science,” *Journal of Science Education and Technology* 22, 6 (2013) 914–926.

²⁰ Tassos A. Mikropoulos, Apostolos Katsikis, Eugenia Nikolou, and Panayiotis Tsakalis, “Virtual Environments in Biology Teaching,” *Journal of Biological Education* 37, 4 (2003): 176–181.

²¹ Melanie Stegman, “Immune attack players perform better on a test of cellular immunology and self confidence than their classmates who play a control video game,” *Faraday Discuss* 169 (2014): 403–423.

²² Carlos Monroy, Yvonne Klisch, and Leslie Miller, “Emerging Contexts for Science Education: Embedding a Forensic Science Game in a Virtual World,” *Proceedings of the 2011 I-Conference: Inspiration, Integrity, and Intrepidity* (New York: Association for Computing Machinery, 2014): 622–629.

²³ Sasha Barab, Tyler Dodge, HakanTuzun, Kirk Job-Sluder, Craig Jackson, Anna Rici, Laura Job-Sluder, Robert Carteaux, Jo Gilbertson, Cohan Heiselt, “The Quest Atlantis Project: A Socially-Responsive Play Space for Learning,” in *The Design and Use of Simulation Computer Games in Education*, eds. Brett Shelton and David Wiley (Rotterdam: Sense Publishers, 2007): 159–186.

For the sake of illustration, let us take a closer look at the first of the programs on this list.

River City was developed by Chris Dede and his colleagues at Harvard University and produced by Acitiv worlds, Inc. It is an immersive virtual reality platform that aims to teach young people about diseases and disease transmission. Users immerse themselves in a fictional 19th century city and learn to behave like scientists. Their aim is to understand why the inhabitants of River City are getting sick and what to do in order to prevent further infection. They do this by ‘talking to various residents in the simulated setting, such as children and adults who have fallen ill, hospital employees, merchants, and university scientists.’ In the process users ‘learn to identify problems through observation and inference, form and test hypotheses, and deduce evidence-based conclusions about underlying causes.’²⁵

Research shows that students who partake in the virtual River City program have much higher rates of success in transferring what they have learnt inside the simulation to the real world. According to Dede:

Our research results from River City show that a broader range of students gain substantial knowledge and skills in scientific inquiry through immersive simulation than through conventional instruction or equivalent learning experiences delivered via a board game. Our findings indicate that students are deeply engaged by this curriculum through actional and symbolic immersion and are developing sophisticated problem-finding skills. Compared with a similar, paper-based curriculum that included laboratory experiences, students overall were more engaged in the immersive interface and learned as much or more.²⁶

Similar results have been found in other studies of educational virtual reality programs.²⁷ The key concept here is that of ‘transfer,’ where a belief or fact learnt inside a *virtual world* is upheld or turned into a belief about the *real world*.²⁸ This

²⁴ Tina Grotzer, Amy Kamarainen, Shari Metcalf, Shane Tutwiler, and Chris Dede, “Teaching the Systems Aspects of Epistemologically Authentic Experimentation in Ecosystems through Immersive Virtual Worlds,” Paper presented at *The National Association of Research in Science Teaching (NARST) Conference*, San Antonio, TX, (April 23, 2017).

²⁵ Chris Dede, “Immersive Interfaces,” 67.

²⁶ Chris Dede, “Immersive Interfaces,” 67.

²⁷ In particular see the studies by Brian Nelson and Diane Ketelhut, “Scientific Inquiry in Educational Multi-user Virtual Environments,” *Educational Psychology Review* 19, 2 (2007): 265–283; Merrilea Mayo, “Video Games: A Route to Large-Scale STEM Education?,” *Science* 323, 5919 (2009): 79–82; Barney Dalgarno and Mark Lee, “What are the learning affordances of 3-D virtual environments?,” *British Journal of Educational Technology* 41, 1 (2010): 10–32.

²⁸ Chris Dede, Jeffrey Jacobson, John Richards, “Introduction,” in *Virtual, Augmented, and Mixed Realities in Education*, eds. Dejian Liu, Chris Dede, Ronghuai Huang, and John Richards (Singapore, Springer, 2017), 6.

suggests that the cognitive process or method involved in forming a belief about the real world includes at least two stages. Firstly, a set of internal cognitive methods are used, such as perception, deduction, induction, etc., in order to arrive at beliefs about the virtual world. Then the participant uses these beliefs to generate a further belief about the real world (see Fig. 1).

It is very likely that beliefs formed inside virtual reality and that are about virtual worlds will have a different meaning or semantic content to beliefs about the real world. Even if I experience objects in a virtual world that are perceptually similar to objects in the real world, such as snow, stop signs, tigers, etc., my beliefs will be about tokens of these 'virtual objects' only, and will not typically include tokens of them in the real world. If I come to believe that 'snow is white' in the virtual world, and on this basis come to form a further belief that 'snow is white' in the real world, then I will have two separate beliefs. The first is a belief about the snow in the virtual world and its property of whiteness, whereas the second is a belief about the snow in the real world and its property of whiteness.

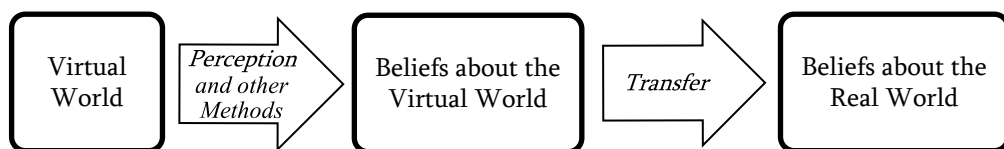


Figure. 1: Virtual World to Real World Belief Transfer

This raises difficult questions such as: 'how do beliefs and statements represent or get to be about virtual worlds?,' 'under what conditions are statements about virtual worlds true?,' and 'what is the ontological status of virtual objects?'. I will not attempt to answer these questions in any detail here. To do so would orientate the discussion too far away from the main epistemological question I want to answer. Instead I refer the interested reader to the ongoing work that is currently being undertaken in this area.²⁹

²⁹ For a pragmatic approach to truth in virtual worlds see Michael Heim, *Virtual Realism* (Oxford: Oxford University Press, 2000). Ilkka Niiniluoto adapts a possible world semantics for fictional worlds and applies it to virtual reality: Ilkka Niiniluoto, "Virtual Worlds, Fiction, and Reality," *Discusiones Filosóficas* 12, 19 (2011):13-28. Theories of the metaphysics of objects in virtual reality include various 'realist' accounts, such as those of David Chalmers, "The Virtual and the Real," *Disputatio* 9, 46 (2017): 309-352; Myeung-Sook Yoh, "The Reality of Virtual Reality," in *Proceedings Seventh International Conference on Virtual Systems and Multimedia*, (Berkeley: CA, 2001): 666-674; Espen Aarseth, "Doors and Perception: Fiction vs. Simulation in Games," *Intermedialités: Histoire et Théorie des Arts, des Lettres et des Techniques* 9, 34 (2007):

Nonetheless, I think there is a very intuitive sense in which a belief or a proposition about a virtual world can be separated from a belief or a proposition about the real world. Literature and other fictional media provide more familiar examples of this. Although the statement 'levitation is possible' is true in the world of *Harry Potter*, this statement is false in the real world. Following David Lewis,³⁰ we might suppose that the difference between them is recognized in ordinary discussion by an implicit prefix of the kind 'In the *Harry Potter* stories, ...' that is attached to the first, but not to the second. Although this is far from a complete analysis of the difference in semantic content between these two statements, it provides a useful way for us to differentiate similar sounding claims about a virtual world from the real world. When an explicit distinction is called for, let us indicate this by using 'P_v' for a statement P that is made about a virtual world, and 'P_r' for a statement P that is made about the real world.

Just like we intuitively recognize a difference between statements about virtual worlds and the real world, we also allow for some of these statements to be true and others false. It is natural, for example, when talking about the video game *Super Mario Bros* to say that 'Mario wears red overalls' is true, whereas the statement 'Mario wears green overalls' is false. Again, how one explains this depends very much on one's theory of truth for virtual worlds and the ontological status of the objects and properties that make these statements true. However, there is one difference between virtual worlds and real worlds that is worth highlighting in connection here. Unlike a work of literature, what is 'true' in a virtual world depends on more than just its program and the intentions of its original creator. It will also depend on the functioning of the hardware that runs the program and the input provided by a user. In addition, there are facts about a virtual world that might lay dormant in its programming because a user did not provide the right input in order for it to manifest. Yet in these cases, we would still want such content to be part of the virtual world, even if it is never actually experienced by a user. If there is a secret level in which Mario's overalls turn

35-44. A phenomenological approach to virtual objects is given in Philip Zhai, *Get Real: A Philosophical Adventure in Virtual Reality* (London: Rowman & Littlefield, 1998). Fictionalist approaches have been developed by Jesper Juul, *Half-Real: Video Games Between Real Rules and Fictional Worlds* (Cambridge, MA: MIT Press, 2005); Cogburn and Silcox, "Against Brain-in-a-Vatism," 561-579. A conditional or dependent realist view (where the reality of virtual objects depends on reproducing qualities of their physical counterparts in the real world) has been developed by Philip Brey, "The Physical and Social Reality of Virtual Worlds," in *The Oxford Handbook of Virtuality*, ed. Mark Grimshaw (Oxford: Oxford University Press, 2014): 42-54.

³⁰ David Lewis, "Truth in Fiction," *American Philosophical Quarterly* 15, 1 (1978): 37-46.

green, we want a statement asserting this fact to be true, even if nobody experiences a 'green Mario.'

Because each implementation or run of the program will produce different states of affairs, it is possible that on different occasions incompatible statements will obtain. For example, for a video game played on one occasion it might be true that 'the boss gets defeated,' whereas on another occasion the statement might be false. Does this imply that there is no consistent conception of truth that can be applied to statements about virtual worlds? I do not believe this observation is necessarily problematic. In these kinds of cases, we might suppose that the language is being used in a rather imprecise way. If I utter the statement 'Manchester United beat Chelsea,' then this is both true and false, since in their histories of competitive soccer, Manchester United has sometimes won against Chelsea and sometimes lost. To be more precise, I would need to qualify the statement to refer to the time or to the actual match played. A similar device could be used in the case of propositions about virtual worlds, where the proposition is suitably indexed to either the time or the run of the program.

3. McBain on Knowledge and Virtual Reality

The case against sensitivity on the basis of virtual reality has been made recently by James McBain. Even if a belief formed in a virtual world is true, the person who holds that belief will not have knowledge because it will fail to be sensitive to the real world:

The person plugged in is not sensitive to that which is true of the situation—that she is being fed stimuli about the world she is experiencing. The method by which she is arriving at her believing this does not counterfactually hold. The details of the world she is experiencing could be changed by the operators. What she is sensitive to is the stimuli, not the world. Therefore, she would not have knowledge in the machine. While the person plugged into the machine will have lots of beliefs about the virtual world she is in, none of those beliefs, on Nozick's account, will constitute knowledge.³¹

If the designers of an educational VR program, such as *River City*, had decided to make the virtual diseases behave in ways quite unlike the real world, then the user would have believed this instead—despite it being false. Their beliefs would not track the truth. Yet intuitively, and empirically, students immersed in the *River City* program can gain knowledge about how diseases function in the real world.

³¹ McBain, "Epistemic Lives," 159.

Whilst staying committed to a broadly externalist epistemology, McBain attempts to explain how knowledge is possible in virtual reality in a way that overcomes the shortcomings of Nozick's sensitivity principle. His solution is to move to an alternative framework, one that includes not counterfactuals, but 'reliable information flow.' Put simply, according to McBain, knowledge in virtual reality is possible provided the beliefs formed are true and are reliably connected to the content of the virtual world. In order to explicate this idea more precisely he adopts Fred Dretske's analysis³² of knowledge as true belief caused by information. There is one important difference, however. Whereas Dretske envisioned the source of information to be a fact or event in the world itself, McBain allows for the source to be a fact or event in the virtual world.

Although McBain's account describes knowledge in terms of Dretske's theory, he suggests it is consistent with Nozick's theory of knowledge and could be reworked with the aid of a modified sensitivity condition.³³ This is what I will attempt to do in this section. First I will explain in more detail how McBain uses Dretske's theory to account for knowledge in virtual reality. Then I will reformulate its central idea in terms of counterfactuals to arrive at a version of sensitivity that can capture the advantages McBain believes his theory has over Nozick's classic tracking approach.

Dretske's original account of knowledge was meant to apply to perceptual belief, and given that perception also plays an important role in knowledge from virtual reality, it seems a suitable place to start. Dretske's central idea is that if a truth-maker for a proposition P transmits the information that 'P' along a channel, such that it is received by S and causes S to believe that P, then S knows that P.

His account relies on a number of key concepts such as 'reliable channel' and 'information'—ideas that are difficult by themselves to understand philosophically. To this end, Dretske utilizes Claude Shannon and Warren Weaver's mathematical theory of communication when thinking about reliability and information flow. According to Shannon and Weaver, information should be measured in terms of the amount of *uncertainty* reduced when a choice is made from a range of possible outcomes.³⁴ If the probability of each outcome has an objective value, then a numerical value can be assigned to indicate how much information each outcome provides. This quantity of information or 'entropy' is measured in terms of the number of binary digits (or bits) needed to individually encode that message.

³² Fred Dretske, *Knowledge and the Flow of Information* (Cambridge, MA: MIT Press, 1981).

³³ McBain, "Epistemic Lives", 162.

³⁴ Claude Shannon and Warren Weaver, *The Mathematical Theory of Communication* (Urbana: The University of Illinois Press, 1949), 1-2.

If one thinks of a source of information x as an object that can be in one of many different states, then a reliable channel can be defined as one that successfully transmits the same *amount of information* along the channel as is generated at the source:

Reliable Channel: In order for a signal r to reliably carry the information that a source x is in state F , the following need to be met:

- (i) x is in state F
- (ii) The signal carries the same amount of information (i.e. in bits) as would be generated if x was in state F
- (iii) The amount of information the signal carries about the source x is or includes the quantity generated by x being in state F (and not by x being in state G)³⁵

Although this tells us when a channel reliably communicates the same amount of information (in bits), it tells us nothing about whether it reliably communicates the same message in terms of its semantic content. If the source is the rolling of a die, then it will transmit the same amount of information whether it lands on a 5 or a 6. Dretske resolves this with the following additional definition:

Semantic Content: A signal r carries the information that ' x is in state F ' if, and only if, the conditional probability of x being in state F given that the signal r transmits the message that ' x is in state F ' is 1 (but less than 1 given the receiver's background beliefs alone).³⁶

This definition claims that a successful transmission of a message occurs only when there is a lawlike connection between the fact itself and the reception of the message. In other words, it is impossible for a channel to transmit the message that ' x is F ' unless x is in state F . As he says 'false information, misinformation, and disinformation are not varieties of information—any more than a decoy duck is a kind of duck.'³⁷ Dretske includes the thought that the probability might be less than 1 given the receiver's background beliefs as a concession to the way we ordinarily talk about information.³⁸ For example, even if the information ' x is in state F ' is true and this is received, it might be not be informative to somebody who already knows it is the case.

Putting all this together, Dretske defines knowledge as follows:

³⁵ Modified from Dretske, "Knowledge," 63-64.

³⁶ Modified from Dretske, "Knowledge," 65.

³⁷ Fred Dretske, "Précis of Knowledge and the Flow of Information," *Behavioral and Brain Sciences* 6, 1 (1983): 57.

³⁸ Dretske, "Précis," 57.

Knowledge: S knows that 'x is in state F' if, and only if, S's belief that 'x is in state F' is caused by the information that 'x is in state F'.³⁹

As an illustration suppose that P is true and P = 'there are 5 apples in the basket.' The world, viewed as a source of information, is in one of many different states and is therefore able to transmit an objective quantity of information. This information is transmitted to the receiver through light waves to the eyes and interpreted to form a belief. If the arrangement of light waves is such that its state mirrors (in terms of number of bits) the state of the world when there are 5 apples in the basket, then it is a reliable channel. If the content of the message is such that it would not have been transmitted unless the odds of it being true are 1, then the right semantic information is also transmitted.

According to McBain, this framework for defining knowledge from ordinary perceptual belief can be extended to virtual reality in the following way:

Once one is hooked up to the machine (or, currently, puts the headset on), the designed world will send signals to the user about that world. The signals will carry as much information about the feature, item, or event being the case in the virtual world as would be generated by that feature, item, or event being the case in the world. Once the user receives the signal, she combines that with any relevant background knowledge about the world, all the while supplementing any gaps with knowledge of the actual world. There is an objective probabilistic connection between what the interface gives and the virtual world being such and such way. If the probability of the virtual world being such and such way when the interface informs us that is 1, then we have grounds for believing the world is that way.⁴⁰

The idea seems to be this: the source of S's belief need not be a fact or event in the real world in order to count as knowledge. Provided the fact or event that occurs in the virtual world carries the same amount of information and the same semantic content as a belief about the real world (and is true), then this can provide knowledge—despite the information not having its origin in the world itself. As McBain puts it: “contra Nozick, [knowledge] is not a matter of where the signals arise.”⁴¹ For example, if S has a perceptual experience of 5 apples in a basket whilst in virtual reality, then S can come to know that 'there are 5 apples in the basket' in the real world provided: (i) there are 5 apples in the basket, (ii) the source of information (i.e. the virtual world) is in a state such that it provides just as much information in bits as 5 apples being in a basket, and (iii) S's belief that

³⁹ Modified from Dretske, “Knowledge,” 65.

⁴⁰ McBain, “Epistemic Lives,” 166.

⁴¹ McBain, “Epistemic Lives,” 167.

'there are 5 apples in the basket' is caused by the information that 'there are 5 apples in the basket' that it received from virtual reality.⁴²

There are a number of things that need to be said about McBain's proposal. Firstly, it is clearly an idealization as it currently stands. Few, if any, existing virtual reality programs can transmit the 'same amount of information' about a virtual object or event as would be received from perceiving it in the real world. Virtual worlds are just not that detailed. Having said that, it could be argued that this a short-term problem, and we can certainly imagine in the future virtual reality software and hardware capable of producing experiences that are as detailed as the real world.

Secondly, we have seen that beliefs about the real world formed on the basis of experiences in virtual reality are likely to have a different semantic content to similar beliefs held about the virtual world itself. Again, this is not necessarily a problem for McBain's proposal. What is needed is to recognize a distinction between the content of the belief about the virtual world and the content of the real world. This does, however, require McBain to restate the supposed connection that obtains between the belief in the virtual world and the real world. It is not enough to say that the belief about the virtual world have 'the same information and semantic content' as a belief about the real world. Even if the worlds are equally detailed (and so contain the same information *vis-à-vis* their entropy) they will necessarily have different content because one is about virtual objects and the other is about real objects. The connection would need to be spelled out in terms other than content identity (perhaps causation or counterfactuals), but I don't see this as being fatal to the proposal.

That McBain's account probably needs supplementing with counterfactuals or a causal connection suggests that it might be worth rephrasing it in terms more familiar to the tracking theory. If we use counterfactual dependence as our mark of reliability in belief formation rather than Dretske's reliable information channel, then McBain's account comes out as follows:

McBain-Inspired Tracking: S knows that P_R only if,

- (i) P_R and P_V are true,

⁴² At times, it reads as if McBain is only concerned to explain how knowledge of the virtual world is possible, rather than knowledge from virtual reality. However, if this is the case then the comparison of his view to Nozick's comments on truth tracking seem ill-placed, as when Nozick rejects knowledge in the machine, he clearly has knowledge of the real world in mind. In the rest of the paper I will use McBain's account as a foundation for explaining knowledge *from* virtual reality, even if McBain did not originally intend it to be used in this way.

(ii) S believes that P_R and S believes that P_V

(iii) If it had not been that case that P_V then S would not have believed that P_V

These conditions capture the core idea latent in McBain's theory: knowledge is attainable from virtual reality provided the user forms a belief about the real world that is true and they base this belief on a belief about the virtual world that is reliably connected to the content of that world. Does this provide an adequate solution to the problem? Can we now say precisely under what circumstances knowledge from virtual reality is attainable? Unfortunately, as I will now argue, the combined conditions (i)-(iii) are neither necessary nor sufficient for knowledge from virtual reality.

4. Problems for Virtual Sensitivity

The proposal given in section 3 on the basis of McBain's account of knowledge effectively weakens Nozick's classic sensitivity in favor of what we might call 'virtual sensitivity.' Instead of a belief formed in VR being sensitive to the facts of the real world (i.e. the facts that make it true), S can have knowledge provided their belief is based on a belief about the virtual world that is sensitive to the facts of the virtual world.

Virtual Sensitivity: Where S bases their belief that P_R on their belief that P_V , if it had not been the case that P_V , then S would not have believed that P_V .

For participants inside a virtual world, clearly this condition is much easier to satisfy than classic sensitivity. A user who is using an educational VR program will continue to believe what they are experiencing even in a world where it is false. But it seems unlikely they would continue to believe it if it were not true in the virtual world itself.

Is virtual sensitivity enough to guarantee knowledge from virtual reality? One worry that immediately emerges is that beyond the condition that P_R is true, there is no reliable method or means connecting S's belief that P_V to the fact that P_R . This raises the logical possibility that a person might come to form a belief about the real world based on one that is true in the virtual world, but that nonetheless is only true by luck. Consider the following case:

COMPUTER MALFUNCTION: A new education VR program has gone to market. In its current form it contains a falsehood about the real world. Whereas P is true in the real world, P is false in the virtual world of the education program. S buys the program and runs it on their computer. However, their computer has a malfunction that incorrectly reads not- P as P , and so when is implemented, creates a visual experience of P . On this basis S comes to believe that P is true, both in the virtual world and the real world.

In this case S has a true belief, both about the virtual world and the real world, and their belief about the virtual world is sensitive to the truths of the virtual world. Yet I would argue that in this case S does not know that P. It might be argued that S is not forming beliefs based on the 'right version' of the VR program, and if their machine had not malfunctioned, their belief would have satisfied virtual sensitivity. But who is to say what the right version is? As we saw in section 2, the virtual world is a combination of many factors: the written program, its implementation on a computing device, and the decisions made by the user. We cannot simply exclude P_v from being true in a virtual world by saying it was not the one intended by the program writer. Many features of virtual worlds were not the result of the intentions of the programmers, often because of complexity and unforeseen consequences. Sometimes this is part of the appeal, but there seems to be no way of excluding this in principle.

It looks as if any proposal along the lines of virtual sensitivity needs to be supplemented with further conditions. It might be thought that it is no surprise that virtual sensitivity (along with true belief) is insufficient for knowledge. After all, Nozick's original tracking theory had two counterfactual conditions: the classic sensitivity condition and his adherence condition. Perhaps the addition of an adherence condition in line with McBain's overall approach is what is called for.

Virtual Adherence: Where S bases their belief that P_R on their belief that P_v , if it had been the case that P_v , then S would not have believed that P_v .

Unfortunately, this does little to change the outcome of the COMPUTER MALFUNCTION case. Here the nearest possible where P_v is true is the actual world, and in this world S believes that P_v . So even though this condition is satisfied, S still does not have knowledge that P_R .

Part of the problem with McBain's proposal (whether couched in terms of counterfactuals or reliable information flow) is that whilst the user's beliefs are sensitive to the facts of the virtual world, they are not sensitive to the real world. This type of sensitivity must be included in the definition to exclude lucky true beliefs based on virtual reality. In the next section I will propose a way in which this can be done that does not require the strength of Nozick's original sensitivity principle.

Before that, however, I want to consider a different case against virtual sensitivity. This case is important for motivating the view I propose in section 5 because it suggests that a user in VR does not need to have knowledge about the virtual world itself:

SPECTRUM INVERSION: A virtual reality program has been designed that deviates systematically from the real world. Every 5 minutes once per hour the

virtual world inverts the colors that are experienced by a user. A user *S* has experienced this world many times and this has caused their brain to compensate for the inverted periods. During an inverted period, the virtual world displays a blue stop sign, however *S* comes to believe that the stop sign in the virtual world is red. On this basis *S* comes to form the belief that stop signs in the real world are also red.

In this case *S* has a true belief about the real world but a false belief about the virtual world. During the period in which they form the belief, the stop sign is blue, and their belief that it is red is false. Yet arguably in this case, *S* *does know* that stop signs are red in the real world. Their belief is connected to the truth in the real world in a non-accidental way. Overall, the medium through which they come to form beliefs (the virtual world) is reliable, and when combined with their own internal compensation for the inverted periods, provides beliefs that are sensitive to the real world.

What this suggests is that even though a belief about the real world may involve, as part of its causal history, an experience and a belief about a virtual world, that belief about the virtual world does not need to be true. In fact, their belief about the virtual world does not even need to be sensitive to the truths of the virtual world. In the SPECTRUM INVERSION case, where P_v = 'stop signs are red,' in the nearest possible world where this is false (the actual world) they believe it. Their belief fails virtual sensitivity and yet somehow they have knowledge of the real world.

The upshot of this example is that although beliefs within a virtual world form part of the causal history for a belief about the real world, their connection to the truth of the virtual world is not significant for knowledge of the real world. These beliefs play a *cognitive* role (they are part of the method of belief-formation) but they do not play an *epistemic* role. In other words, knowledge of the real world based on an experience in virtual reality does not first require a person to have knowledge about the virtual world. Perhaps this should not come as a surprise. Most externalist views of perception and testimony, for example, only require that a person's belief be connected to the facts in the right way; they do not demand the stronger requirement that they first have beliefs or knowledge about the various stages in the generation of their belief. Having a true belief about the virtual world is clearly important for knowledge about the virtual world, but when thinking about using virtual reality to gain knowledge about the real world, the truth of these beliefs no longer seems necessary.

5. Fixing the Problem: Virtual Sensitivity+

In this final section I will introduce a new version of sensitivity that is stronger than the McBain-inspired virtual sensitivity principle, but weaker than Nozick's original classic sensitivity principle. The motivating idea will be that where a belief about the real world is based on one about the virtual world, a person can come to know about the real world provided there is a non-accidental, lawlike, connection between the facts of the real world and the beliefs they form inside the virtual world. In other words, their beliefs about the virtual world are sensitive to the real world.

This idea allows us to separate brain-in-a-vat scenarios where an envatted person can have knowledge from those in which they cannot. Consider the following two cases:

EVIL SCIENTIST: S has been envatted her entire life and is fed sensory stimuli by an evil scientist. The evil scientist uses all kinds of methods to decide on the content of the world experienced by S. One day the scientist uses a random number generator to decide how many fingers S's avatar will have in the virtual world. The outcome of the random number generator is 5 and S's avatar in the virtual world comes to have 5 fingers.

BENEVOLENT SCIENTIST: S has been envatted her entire life and is fed sensory stimuli by a benevolent scientist. The benevolent scientist uses the best methods they can to ensure that the content of the world experienced by S is as close as possible to the real world. Knowing that humans have 5 fingers, the scientist programs the virtual world so that S's avatar has 5 fingers.

Now let us suppose that, for one reason or another, S is given a human body and returns to the real world. Many of the beliefs she formed whilst a brain in a vat are transferred to beliefs about the real world. Do any of these constitute knowledge? A good case can be made for arguing that beliefs formed in the EVIL SCIENTIST scenario will not count as knowledge (even if true), whereas those formed in the BENEVOLENT SCIENTIST scenario will. In the BENEVOLENT SCIENTIST case, if it had been true that humans had 6 fingers, then it is probable that S would have believed this about her avatar's hands instead. In his attempt to make her experiences as close to the real world as possible, it is likely that the benevolent scientist would have included this fact in the virtual world and that S would have come to believe it.

The sensitivity that is called for here is one that is stronger than virtual sensitivity. For that reason, let us call it 'virtual sensitivity+':

Virtual Sensitivity+: Where S bases their belief that P_R on their belief that P_V , if it had not been the case that P_R , then S would not have believed that P_V .

Putting it all together, on the proposal being suggested here, S can come to know that P (is true about the real world) provided P is true, S believes that P and S's belief that P meets virtual sensitivity+.

We can now explain why the individual in the COMPUTER MALFUNCTION case does not have knowledge. Even though their belief about the real world is true and is based on a belief about the virtual world, their belief in the virtual world is not sensitive to the truth of the real world. In this scenario, let us suppose that P is false. In which case the design of the educational VR program would now accurately reflect the truth of the real world. However, once again there is a malfunction and instead of the machine producing a virtual world where P is false, it produces one in which P is true. S comes to form the belief that P based on their experiences. But now we can see what went wrong. Here their belief is not sensitive to the truth of the real world: in the nearest possible world where P is false, S would continue to believe that P is true. The belief formed in COMPUTER MALFUNCTION fails virtual sensitivity+ and therefore does not amount to knowledge.

What about the SPECTRUM INVERSION case? This example suggests that beliefs about a virtual world do not need to be true in order to provide knowledge of the real world. At first glance this looks puzzling and even counterintuitive. But if virtual sensitivity+ provides the right modal relationship between the facts and a person's belief, then we can explain why this is the case. In SPECTRUM INVERSION the individual's belief is connected via a reliable mechanism to the facts of the real world. That mechanism involves a number of steps: (i) the content of the virtual world has been designed intentionally to match the real world, (ii) the virtual world inverts periodically, (iii) S's perceptual and belief-forming mechanism compensates for the inversion stage. The result is that when a belief is generated at the end of this chain, it is sensitive to the facts of the real world.

To demonstrate this last point, let us consider what would happen in the nearby worlds where P is false, i.e. where stop signs are not red. Let us suppose that they are blue. Then the virtual world created would invert colours for 5 minutes every hour. S happens to experience the colour of the virtual stop sign during these 5 minutes, which whilst inverted, is actually red. Because their perceptual and cognitive faculties have compensated for this fact, they form the belief that stop signs are blue in the virtual world, and therefore, that stop signs are blue in the real world. In the nearby world where P is false in the real world, they also come to believe that it is false in the virtual world. Therefore, virtual sensitivity+ is satisfied and this fits with our intuitions that in this case the person would have attained knowledge.

6. Conclusion

I have argued that beliefs can amount to knowledge even if the sole basis for believing them comes from virtual reality. This outcome is the one that is most consistent with the empirical evidence surrounding the role of virtual reality and related technologies in training and education. Accommodating this knowledge within Nozick's truth-tracking framework requires modifying the sensitivity principle to fit the unique stages of belief-formation present when a belief from a virtual world is transferred to the real world. Given that bespoke versions of the sensitivity principle have been shown to be necessary for other methods of belief-formation, this is not a unique problem for virtual reality. Ultimately, philosophers need to pay more attention to the epistemic aspects of virtual reality, which, as the examples highlighted above demonstrate, is likely to play a much greater role in the formation of our beliefs in the future.⁴³

⁴³ Acknowledgements: I am grateful to the participants of the 7th National Meeting in Analytic Philosophy (Lisbon Portugal, 09/2018) for helpful comments and constructive criticism on an earlier draft of this paper. I also wish to thank the Centre for Ethics, Politics and Society at the University of Minho (Braga Portugal, 09/2018) who invited me to present this paper as part of their guest seminar series.

DISCUSSION NOTES/ DEBATE

DOES METAPHILOSOPHICALLY PRAGMATIST ANTI-SKEPTICISM WORK?

Scott AIKIN

ABSTRACT: Michael Hannon has recently given “a new *apraxia*” argument against skepticism. Hannon’s case is that skepticism depends on a theory of knowledge that makes the concept “useless and uninteresting.” Three arguments rebutting Hannon’s metaphilosophical pragmatism are given that show that the concept of knowledge that makes skepticism plausible is both interesting and useful.

KEYWORDS: skepticism, anti-skepticism, pragmatism, metaphilosophy

1.

Anti-skepticism comes in four flavors. There is (1) the quest of directly answering the skeptical challenges, (2) the program of showing that skepticism is self-defeating, (3) the line that the skeptic has artificially set the standards for knowledge (or other relevant epistemic property) too high, and then there is (4) the argument that skepticism yields objectionable practical results. Call these the *heroic*, *self-refutative*, *redefinitive*, and *pragmatic* arguments, respectively.¹ Of the pragmatic arguments against skepticism, there is a prominent subset that are best termed *apraxia* arguments – that were we to believe skepticism is true, then we would not be able to get on with our lives. The primary targets for the *apraxia* argument are skepticisms that require suspension of belief in light of the fact that few items of reflection survive skeptical scrutiny. Given that intentional action requires belief, skepticism stands in the way of one living one’s life. The skeptic, so the argument goes, is paralyzed.

The problem with the *apraxia* objection, as should be clear when stated so starkly, is that it does not follow that skepticism is *false* if it is *inconvenient*. In short: that the fact that a philosophical view that has bad practical consequences is not sufficient evidence that the view is false, but only that we should prefer it so.

¹ See Scott Aikin and Thomas Dabay, “Pragmatist Anti-Skepticism: At What Cost?” in *The Mystery of Skepticism*, eds. Kevin McCain and Ted Poston (Leiden: Brill, 2018) for a short overview of these. Hannon invokes these four approvingly to locate his “new *apraxia* argument.”

Call this *the fallacy of inconvenience*. The *apraxia* argument, then, is better a *motivation for developing* anti-skeptical arguments of other stripes than itself being one.

Michael Hannon, in a recent essay, “Skepticism: Impractical, therefore Implausible,” announces that he will argue for “the *new apraxia* objection,” one that targets not the impractical results of the skeptic’s theoretical view, but rather the skeptic’s theoretical position as one that “goes against the very purpose of theoretical evaluation.”² His thesis is that apraxia arguments show that the skeptics’ views undercut the point of our concept of knowledge. And notice how the view, framed as such, does not yield the fallacy of inconvenience, since the practical edge of the argument is about how the concept is defined, not whether its applications are convenient or not. It is about *the point of the concept* of knowledge. Thus, a *metaphilosophically pragmatist anti-skepticism* – it is *pragmatist* anti-skepticism because it is a version of the *apraxia* argument, and it is *metaphilosophical* because it begins with a view about the point of philosophical reflections on knowledge.

I will argue here that Hannon’s argument has three complications, and I think that these complications should give us pause with the pragmatist anti-skeptical program. Instead, I think, what Hannon’s argument shows is something that skeptics have thought for a long time – that our epistemic concepts have a variety of equally plausible but inconsistent valences. Skepticism, then, isn’t just a first-order view about knowledge, but it’s a view about our views of knowledge, too. And so, a metaphilosophical skeptical defense of skepticism is in order. To close, I will outline reasons why the concept of knowledge behind skepticism is worth having.

2.

In order to avoid committing the fallacy of inconvenience, Hannon designs his argument to target some desiderata for a theory of knowledge. The thought is that if it can be shown that a theory fails some requirements of what we would hope for with a theory, we’ve shown that it fails as a theory. Those purposes are what Hannon calls “adequacy conditions” on a theory of knowledge. The basic structure of the argument works as follows:

1. A theory of knowledge is adequate only if it fits plausible assumptions about the point of having the concept of knowledge.
2. Skepticism does not fit plausible assumptions about the point of having the

² Michael Hannon, “Skepticism: Impractical, Therefore Implausible,” *Philosophical Issues* 21, 1 (2019): 143-158.

concept of knowledge.

Therefore, skepticism is not an adequate theory of knowledge.

Hannon frames the argument as follows:

[S]kepticism is unable to underwrite the primary roles that our knowledge concept plays in epistemic evaluation. This is because skepticism has no connection to the practical circumstances that explain why we speak of knowing in the first place. On these grounds [...] skepticism should be regarded as an implausible theory of knowledge.³

The two crucial elements of Hannon's argument are (i) making the case for the adequacy conditions, and (ii) showing that skepticism fails them. Hannon holds that there are four adequacy conditions connected to what might be called *the point of the concept of knowledge*.

The first is that "the primary function" of the concept of knowledge is "to *identify reliable informants*."⁴ We use the concept *knowledge* and *knower* to distinguish those on whom we should rely from those who we should not. So, just as it would be silly to make it *too easy* to qualify as a knower, it would be equally pointless to make it *too hard* for people to qualify, too. The skeptical result with knowledge, as Hannon puts it, "runs against this [social-epistemological] approach, because it would frustrate our communal epistemic practices." If skepticism is right, then this social sorting point of the knowledge concept is frustrated, and "we have no use for such a concept."⁵

Hannon's second desideratum of a theory of knowledge is what he calls its "inquiry-stopping function." The basic thought is that upon meeting the conditions for knowledge, we may responsibly stop inquiring and get on with what we were doing. The concept of knowledge is useful because it serves as the limit for when inquiry has gone far enough. Hannon's reasoning, then, invokes a pragmatist point about inquiry:

[T]o continue to inquire beyond a certain point would be impractical: it would commit us to paying higher informational costs that are worth the lessened risk of being wrong.⁶

³ Hannon, "Skepticism: Impractical, Therefore Implausible," 145.

⁴ Hannon, "Skepticism: Impractical, Therefore Implausible," 146.

⁵ Hannon, "Skepticism: Impractical, Therefore Implausible," 5.

⁶ Hannon, "Skepticism: Impractical, Therefore Implausible," 148. See also Michael Hannon, *What's the Point of Knowledge?* (Oxford: Oxford University Press, 2019), 215.

In short, Hannon's objection is that the skeptical theory of knowledge is bad economics with our limited epistemic resources – we, given the skeptic's program, would continue inquiry past all reasonable limits.

Hannon's third and fourth arguments are connected, as they are manifestations of the familiar *knowledge norms* – the norm of assertion and the norm of action. Hannon captures the norm of assertion as follows:

Suppose, first, that you are in a good enough epistemic position to assert that p if (and perhaps only if) you know that p. If skepticism were true, then assertions could almost never be epistemically warranted [...] But such a result seems obviously intolerable from a practical standpoint, because we often have urgent needs for communicating information.⁷

The norm of action has a similar role:

[A]ssume that you are in a good position to rely on p in practical reasoning if (and perhaps only if) you know that p. If skepticism were true, your justification would almost never provide a sufficient basis for practical reasoning.

The problem, as Hannon puts it, is “we are still faced with the unavoidable need to act,” and so “there is practical pressure to think knowledge is the relevant norm only if skepticism is false.”⁸ The result is that with both assertions and actions consequent of practical reasoning, skepticism renders those activities “impossible unless we constantly violated the epistemic norms governing those practices.”⁹ The skeptic, in short, would have a theory of our practices that makes it so that we never properly practice them. The problem, then, is in what sense these would be *practices* at all? The turn, then, is that skepticism's theory of knowledge is unsupported – it's not clear what it would be a theory of.¹⁰

Hannon's overall argument is that given these four convergent arguments, we have reason to be committed to the claim that skepticism is “implausible because it is impractical,” since “it goes against the point of epistemic evaluation.”¹¹ That is, the conception of knowledge that would make skepticism look plausible is a conception of knowledge that has no purpose. As Hannon frames it, “Either the purpose of knowledge is such that it rules out skepticism or else knowledge is a useless and uninteresting concept.”¹²

⁷ Hannon, “Skepticism: Impractical, Therefore Implausible,” 149.

⁸ Hannon, “Skepticism: Impractical, Therefore Implausible,” 149.

⁹ Hannon, “Skepticism: Impractical, Therefore Implausible,” 149.

¹⁰ Hannon, *What's the Point of Knowledge?*, 218.

¹¹ Hannon, “Skepticism: Impractical, Therefore Implausible,” 156.

¹² Hannon, “Skepticism: Impractical, Therefore Implausible,” 149.

3.

It should be clear why Hannon's classifies his argument as a form of *pragmatist* anti-skepticism, and I believe it should be clear why it is, in the end, a *metaphilosophically pragmatist* argument, too – it is about *the point* of our concepts and what our theories about those concepts must be in the service of, namely, effective practice. The anti-skeptical challenge amounts to asking why we would conceive knowledge in such a way that would make skepticism plausible. Hannon's four cases are reasons to conceive knowledge otherwise.

I believe there are three metaphilosophical defenses for the skeptic here, and the first begins with an historical explanation. Skepticism, classically, was not a free-standing philosophical tradition – it was one that was a critical reply to the 'dogmatic' philosophical traditions around it. Academic and Pyrrhonian skepticisms were internal critiques of Stoic and Epicurean epistemology.¹³ The Stoics required that knowledge begin with *kataleptic* impressions, which had not only to be true, but they had to be caused by what they represented and could not be confused with false impressions. And so, it was from Stoic epistemology that skeptics found reason to propose indiscernibility cases. And the same goes for the Epicureans. They held that all sensations are true, and so they were ripe for the problem of perceptual variance. But this point generalizes – philosophical theories that require significant revision to how we live, what we think of ourselves and the world, and how we conceive of the good must propose accounts of how these things are known. And in particular, they must explain further how we know these things when others do not. Revisionary philosophical programs then require high-grade standards for knowledge, otherwise they cannot explain why we should follow their dictates instead of those of our unenlightened fellows.

So, the historical point is that the theory of knowledge the skeptics use had itself been derived from going non-skeptical epistemologies. And there is a reason to have such high-grade requirements – reflection on what knowledge is (and other core concepts to our lives, such as beauty, the good, reality, and justice) requires that high-grade requirement, as when that requirement is satisfied, it provides powerful reason for changing our lives for the better. And so, I believe, the historical explanation yields the first response favoring the skeptics – the clarification of the concept of knowledge allows us the tools to identify things we can be confident in, things worth changing our lives in light of. And so, this historical defense is not just a defense of skepticism in particular, but it is a defense

¹³ See Scott Aikin, "Skeptics against Epicureans and Stoics on the Criterion," in *The Routledge Handbook for Hellenistic Philosophy*, ed. Kelly Arenson (New York: Routledge, 2020) 191-203.

of any program of significant revision of our concepts and re-orienting of our lives. You don't need merely what passes for knowledge, but the high-grade stuff that *really is knowledge*.

The skeptics deny that we have any of that high-grade knowledge stuff, but it's not out of their desire to be obtuse that they say so. Rather, it's out of the pursuit and complications in the pursuit of that epistemic good that they say it. It is an unhappy result, but it is not one that is simply *pointless*.

The second skeptical defense falls hard on the heels of the first historical line of argument. It opens with the questions: Are we so sure what the point of our concept of knowledge is, to begin with? Is the point one that entails that there are instances? Notice that all of Hannon's four cases require that the point of the concept entails that there are instances that get sorted as successful. Hannon's challenge is that if the concept of knowledge does not yield instances in these four domains, then it is "pointless" or "uninteresting." But notice that there are many concepts that have an aspirational edge that have no guarantees of instances. Take *justice* for example. Imagine someone to have a theory of justice that is demanding, perhaps so demanding that there are no states or laws that, at the end of analysis, satisfy its requirements. So it follows that, on this theory of justice, there are no just states or laws. If this were a well-motivated theory, this result, I think, would be *supremely interesting*. Ask any philosophical anarchist. Or consider a theory of what it takes for something to be morally good. If, again, the theory were well-motivated but yielded a nihilism of good actions and agents, that would be, again, supremely interesting. Ask anyone who asks critical questions about moral saints. And we can do this with other simple notions like *scholarly duties*, *parental care*, and *teacherly excellence* – there seems to be no upper limit on what we can do in their service to perform them. Surely it is useful to have a theory that captures that notion that our tasks are incomplete, even when we've done our right best. We may be blameless for leaving off, but that does not mean we've satisfied the demands of the task. The same, as I see it, can be said of knowledge, too. And not only, I believe, is it *interesting*, but it's *useful*.

One way to capture the usefulness of these concepts is to turn back to Hannon's fallibilist alternative to skepticism. Now, the skeptics, too, were fallibilists (particularly the Academics) – but fallibilists about *reasonable belief*, not knowledge. Academics distinguished between three levels of worthiness of assent for impressions: (a) plausible, (b) plausible and tested, and (c) plausible, tested, and stable. In these various instances, we can more reasonably assent, act, take as reliable, and (temporarily) discontinue inquiry, but we don't need to concept of

knowledge to do so.¹⁴ The lesson is that we need only these degreed notions of reasonable belief to perform these acts. We need the concept of knowledge to explain why these cases aren't always right – namely, that though we had good reasons to do what we did, we nevertheless didn't *know*. And we have a way to explain why we do not just close inquiry, but that we re-open it – namely, that though we had good reasons, we nevertheless did not know. The concept of knowledge, then, plays a regulative role on our notions of responsible practice for the skeptics, and it does have a purpose.

The third and final metaphilosophical defense of skepticism is simply from the following counterfactual. *Skeptical challenges would not be so easily posed if skeptics used an alien or confabulated concept of knowledge*. The regress problem is posed by five-year-olds, but it's the *anti-skeptics* that have to do the fancy philosophical footwork to say what went wrong (and, by the way, they don't say all the same thing!). Disagreement skeptics need only the notion of *has the same evidence* to pose their challenge, but it's the anti-skeptics that have to say complicated (and sometimes pretty dogmatic) things to avoid skeptical results. Not one viewer of *The Matrix* had to be taught the closure principle to wonder if they, too, were in the Matrix. Who is doing the conceptual re-engineering here? I note all of this to highlight the fact that the high-grade concept of knowledge, the one that makes skepticism possible but also the idea that we can have profound insight, too, is a useful, interesting, and familiar notion. It allows us to, even when we are very sure, to state our lingering doubts, it keeps us intellectually humble, and it drives us to improve. That ain't nothing.

4.

Michael Hannon has given what he calls 'the new *apraxia* argument' against skepticism. Hannon's core thesis is that a theory of knowledge must not run afoul of why the concept of knowledge is useful, and he outlines four desiderata. They are that the concept of knowledge is (1) for identifying reliable cognitive resources, (2) for closing inquiry, and for (3) asserting and (4) practical reasoning. Since skepticism is the view that there are no instances of knowledge, the concept of knowledge is rendered without use. Since we nevertheless *do* perform these actions, we should reorient the concept of knowledge. Hannon's metaphilosophical pragmatist program can be answered by three metaphilosophical arguments in favor of the concept of knowledge that makes skepticism plausible. The first is that

¹⁴ See Sextus Empiricus's account of Academic skeptical fallibilism in *Outlines of Pyrrhonism* I.227-230 and Cicero's at *Academica* 2.66-8.

Scott Aikin

skeptics use the concept of knowledge that drives any revisionary philosophical program – we need an epistemically high-grade notion of knowledge to explain the basis on which we reorient ourselves. Second, all four of Hannon's desiderata for the concept of knowledge can be, given Academic fallibilism, handled by the gradable notion of reasonable belief, and the concept of knowledge instead plays a regulative role over those functions. Third, and finally, there is clear evidence that the concept of knowledge that makes skepticism plausible is familiar and considerably less controversial than the going products of anti-skeptical re-engineering programs. This, of course, is not an argument for skepticism, but rather a rebutting case against the pragmatist case for throwing out the high-grade notion of knowledge that makes skepticism plausible.

NOTES ON THE CONTRIBUTORS

Scott F. Aikin is Associate Professor and Director of Undergraduate Studies in Philosophy at Vanderbilt University. He specializes in epistemology, argumentation theory, and ancient philosophy. He is the author of *Epistemology and the Regress Problem* (Routledge, 2011), *Evidentialism and the Will to Believe* (Bloomsbury, 2014), and *Political Argument in a Polarized Age* (with Robert Talisse, Polity, 2020). Contact: scott.f.aikin@vanderbilt.edu.

Robb Dunphy is a Junior Research Fellow at the Maimonides Centre for Advanced Studies at the University of Hamburg. He previously taught philosophy at the University of Sussex. His research is especially concentrated on Pyrrhonian scepticism, German Idealism, and epistemology. His published work has appeared in journals including the *Review of Metaphysics*, *Hegel Bulletin*, and *Studies in Social and Political Thought*. Contact: robbdunphy@gmail.com.

Domingos Faria holds a FCT junior researcher position (CEECIND/01066/2017) at the Center of Philosophy of the University of Lisbon (CFUL), where he develops a project entitled “The Epistemology of Complex Disagreement Among Social Groups” at the LanCog group. Affiliation: LanCog, Centre of Philosophy, University of Lisbon. Contact: domingosfaria@campus.ul.pt.

Thomas Grundmann is professor of philosophy at the Department of Philosophy, University of Cologne. Much of his research focuses on themes from general epistemology, a priori knowledge, philosophical methodology, and social epistemology. His most recent research projects concern the epistemology of disagreement, the epistemology of fake news and post truth, and epistemic authority. Some relevant publications include: “How to Respond Rationally to Peer Disagreement: The Preemption View” (*Philosophical Issues*, forthcoming), “Why Disagreement-Based Skepticism Cannot Escape the Challenge of Self-Defeat” (*Episteme*, forthcoming), “Doubts about Philosophy? The Alleged Challenge from Disagreement” (in *Knowledge, Virtue, and Action: Putting Epistemic Virtues to Work*, 2013). Contact: thomas.grundmann@uni-koeln.de.

Perry Hendricks is a PhD student at Purdue University. His primary interests are epistemology, philosophy of religion, and bioethics. He has published articles in

such journals as *Pacific Philosophical Quarterly*, *The Journal of the American Philosophical Association*, and *Bioethics*. Contact: ampchendricks@gmail.com.

Robert Michels studied philosophy in Constance, Germany and Geneva, Switzerland, where he also received his PhD after defending a thesis on the Essentialist Theory of Modality written under the supervision of Kevin Mulligan. He is currently a postdoc at the Università della Svizzera italiana within the SNSF-funded project “Identity in Cognitive Science, Quantum Mechanics, and Metaphysics.” He works on a range of topics which are mostly connected to modality or indeterminacy in metaphysics, philosophy of language, epistemology, general philosophy of science, and the history of analytic philosophy. Website: <http://www.robert-michels.de>. Contact: robert.michels@usi.ch.

Diana Sofronieva obtained her PhD from the University of Leeds, UK. Her primary research interests concern empathy, moral character, and friendship. You can find her online at sofronieva.com. Contact: hello@sofronieva.com.

Billy Wheeler is Research Fellow in the Department of Philosophy (Zhuhai) at Sun Yat-Sen University. He completed his PhD in the Department of History and Philosophy of Science at the University of Cambridge in 2013. Billy is the author of the book *Idealization and the Laws of Nature* (Springer, 2018) and has published articles in the journals *Theoria*, *Filozofia Nauki*, *Manuscrito*, *Principia*, and *Techné*. His main work investigates the epistemological, metaphysical and ethical implications of emerging technologies. He has particular interests in virtual reality, robotics and big data. Contact: billy.wheeler@cantab.net.

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, French and German, 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.