

TABLE OF CONTENTS

Articles

Scott F. AIKIN, Michael HARBOUR, Jonathan NEUFELD, Robert B. TALISSE, Epistemic Abstainers, Epistemic Martyrs, and Epistemic Converts.....	211
John A. BARKER, Fred ADAMS, Epistemic Closure and Skepticism.....	221
Vihren BOUZOV, Scientific Rationality as Normative System.....	247
James CARGILE, Two Fallacies.....	257
Murray CLARKE, Concepts, Intuitions and Epistemic Norms.....	269
Sorin COSTREIE, Frege's Context Principle: its Role and Interpretation....	287
Theodore J. EVERETT, Observation and Induction.....	303
Carl GINET, Self-Evidence.....	325
Steven D. HALES, No Time Travel for Presentists.....	353
Lauren J. LEYDON-HARDY, Getting Gettier'd on Testimony.....	361

Reviews

Mark R. WYNN, <i>Faith and Place. An Essay in Embodied Religious Epistemology</i> , reviewed by Ioan Alexandru Tofan.....	373
Cory JUHL, Eric LOOMIS, <i>Analyticity</i> , reviewed by Viorel Țuțui.....	379
Notes on the Contributors	387
<i>Logos & Episteme</i> . Aims and Scope	391
Notes to Contributors	393

ARTICLES

EPISTEMIC ABSTAINERS, EPISTEMIC MARTYRS, AND EPISTEMIC CONVERTS

Scott F. AIKIN, Michael HARBOUR,
Jonathan NEUFELD, Robert B. TALISSE

ABSTRACT: An intuitive view regarding the epistemic significance of disagreement says that when epistemic peers disagree, they should suspend judgment. This abstemious view seems to embody a kind of detachment appropriate for rational beings; moreover, it seems to promote a kind of conciliatory inclination that makes for irenic and cooperative further discussion. Like many strategies for cooperation, however, the abstemious view creates opportunities for free-riding. In this essay, the authors argue that the believer who suspends judgment in the face of peer disagreement is vulnerable to a kind of manipulation on the part of more tenacious peers. The result is that the abstemious view can have the effect of encouraging dogmatism.

KEYWORDS: epistemology, disagreement, dialogue

Two people, Alf and Betty, disagree. Alf believes that *p*, and Betty believes that not-*p*. Suppose that Alf and Betty are *epistemic peers*—they share roughly the same evidence and neither is more intellectually capable than the other in any substantive way. Also, allow that Alf and Betty have discussed thoroughly each other's reasons, but neither has been moved.

This circumstance is troubling for Alf and Betty, and not just for *practical* purposes (perhaps they'd like to resolve their disagreement because they have plans that depend upon their agreement with respect to *p*) but also for *epistemic* purposes. For each, the fact that an epistemic peer disagrees calls into question the quality of the reasons supporting their respective beliefs. Even though neither can say precisely where the other has gone wrong, they nevertheless each hold that the other's case does not yield reason to justify the other's belief.

Richard Feldman¹ has argued for the intuitive view that under circumstances where one is an apparently reasonable believer and affirms a proposition that an epistemic peer denies, one should *suspend judgment*:

One of us must be making some kind of mistake or failing to see some truth. But I have no basis for thinking that the one making the mistake is him rather than me. And the same is true of him. And in that case, the right thing for both of us to do is suspend judgment on *P*².

According to Feldman, although there *seem* to be reasonable disagreements among epistemic peers—viz., disagreements where each party is within his or her epistemic rights to hold his or her respective view—this is in fact an illusion. Consequently, in cases where it appears that one is reasonably disagreeing with a peer, one “should suspend judgment about the matter under dispute.”³ Hence Feldman holds the *principle of suspension*:

(PS) If S disagrees with an epistemic peer about *p*, then S should suspend judgment about *p*.

The case for PS depends on a principle regulating evidence which Feldman calls *The Uniqueness Thesis*:

(UT) A body of evidence justifies at most one proposition of a competing set of propositions and... it justifies at most one attitude toward any proposition.⁴

The rationale for UT is, we think, also intuitive: A body of evidence either supports *p* or it does not. And if it does, one is justified in believing that *p* on the basis of that evidence. But if not, one is either justified in believing not-*p* (or some specific competing proposition exclusive of *p* supported by the evidence), or one should suspend judgment with regard to *p*. Feldman takes peer disagreement to place believers under an obligation to justify their preference for their own belief over their peers; consequently, disagreement among peers gives rise to extra epistemic burdens. And in cases of disagreement among peers, believers have no

¹ Richard Feldman, “Reasonable Religious Disagreements,” in *Philosophers Without Gods: Meditations on Atheism and the Secular Life*, ed. Louise Antony (New York: Oxford University Press, 2007), 194–214, and “Evidentialism, Higher-Order Evidence, and Disagreement,” *Episteme* 6 (2009): 294–312.

² Feldman, “Reasonable,” 212.

³ Feldman, “Reasonable,” 212.

⁴ Feldman, “Reasonable,” 205.

non-question-begging way to provide the required justificatory story since, by hypothesis, the disagreeing peers share all the same evidence and are equally capable cognitive agents. Consequently, Feldman holds that “a peer that disagrees with you is evidence against the view you believe.”⁵ Thus, Feldman concludes, when faced with a disagreeing peer, one must suspend judgment.

The view Feldman espouses here is intuitively attractive, and is a member of a broad family of views regarding the epistemology of disagreement one may characterize as *epistemic abstemiousness*. The abstemious view is roughly that if one finds oneself in disagreement with another that is ostensibly neither better nor less informed on the issue, then one has evidence of equal weight between two inconsistent propositions. Consequently, one should abstain from belief—that is, suspend judgment. In precisely this idiom, Sextus Empiricus took disagreement to be its own autonomous skeptical mode:

According to the mode deriving from dispute, we (the skeptics) find that undecidable dissention about the matter proposed as come about both in ordinary life and among philosophers. Because of this we are not able either to choose or rule out anything, and we end up with suspension of judgment⁶

Versions of epistemic abstemiousness are commonly found in discussion of theological diversity in the philosophy of religion. For example, William Cantwell Smith has argued on moral grounds that one has a cognitive duty of intellectual humility:

[E]xcept at the cost of insensitivity or delinquency, it is morally not possible actually to go out into the world and say to devout, intelligent, fellow human beings: ‘... we believe we know God, and we are right; you believe you know God, and you are wrong.’⁷

John Hick has argued similarly:

⁵ Feldman, “Evidentialism,” 331.

⁶ Sextus Empiricus, *Outlines of Scepticism*, trans. Julia Annas and Jonathan Barnes (New York: Cambridge University Press, 2000), PH I.165.

⁷ Wilfred Cantwell-Smith, *Religious Diversity* (New York: Harper and Row, 1976), 14.

Nor can we reasonably claim that our own form of religious experience, together with that of the tradition of which we are a part, is veridical whilst others are not.⁸

And Gary Gutting contends that once one has been made aware of disagreement regarding a religious belief, one's epistemic duties increase—one must not only be justified in believing, but one must be able to account for the fact of the disagreement. According to Gutting, in the face of unresolved disagreement, one must dampen one's commitment: one must withdraw *decisive assent* and extend to one's belief only *interim assent*.⁹

Abstemiousness has been recommended by philosophers working outside of philosophy of religion as well. Keith Lehrer has argued that disagreements between genuine inquirers are rationally impossible:

Actual disagreement among experts must result either from an incomplete exchange of information, individual dogmatism, or a failure to grasp the mathematical implications of their initial state and yet disagree.¹⁰

Crispin Wright's account of *cognitive command* entails a similar result. A discourse has cognitive command when, if given differing opinions on a matter in the discourse, one knows *a priori* the divergence must be explainable in terms of at least one of the views having an *imperfection of pedigree*. Accordingly, when cognitive command is present a "cognitive shortcoming *always* has to be at work in the generation of conflicting views."¹¹

In contemporary discussions of the epistemology of disagreement, the connection between what Bogardus¹² and others¹³ have called the *equal weight view* and the conciliatory inclinations we see in abstemiousness is widely recognized. Christensen holds that in cases of peer disagreement one should often

⁸ John Hick, *An Interpretation of Religion*, 2nd Edition (New Haven: Yale University Press, 2004), 235.

⁹ Gary Gutting, *Religious Belief and Religious Skepticism* (Notre Dame: Notre Dame University Press, 1982), 105.

¹⁰ Keith Lehrer, "When Rational Disagreement is Impossible," *Nous* 10 (1976): 331.

¹¹ Crispin Wright, *Truth and Objectivity* (Cambridge: Harvard University Press, 1992), 147; Cf. Crispin Wright, "On Being in a Quandary. Relativism vagueness logical revisionism," *Mind* 110 (2001): 58.

¹² Tomas Bogardus, "A Vindication of the Equal Weight View," *Episteme* 6 (2009): 324-335.

¹³ Adam Elga, "Reflection and Disagreement," *Nous* 41 (2007): 478-502, Roger White, "On Treating Oneself and Others as Thermometers," *Episteme* 6 (2009): 233-50.

“split the difference.”¹⁴ Earl Conee holds that when faced with a mature dispute between mutually recognized epistemic peers, even when one is one of the disputants, one must take a third-person view of the situation and reason:

Either way, we are justified in thinking that minds having equally good prospects of finding out the truth about X are on each side. In light of this, the reasons we have been given for and against X remain in balance.¹⁵

This is because, from the disinterested view, “we have no better basis for discounting opposing summary impressions than we do for our own.”¹⁶

As we have said, it is intuitive to think that when we find ourselves disagreeing with an epistemic peer, we must abstain from belief. The abstemious view seems to capture a kind of detachment requisite for rational beings and a kind of conciliatory inclination that makes for irenic discussion. Feldman’s articulation of the abstemious view seems to us the most explicit version yet proposed. But is it correct? We think not. In fact, we shall argue that, despite its initial intuitive appeal, the abstemious view yields results that are highly counterintuitive.

Return to our peers who disagree, Alf and Betty. Assume Betty accepts Feldman’s PS and so reasons as follows:

Because Alf is an epistemic peer who disagrees with me with respect to p, I must suspend judgment with respect to p.

And then Betty suspends. But note that this introduces a drastic shift in the epistemic situation that obtains between Alf and Betty. Betty originally held that not-p, but *now*, just with a bit of discussion, Betty has weakened her belief that not-p to suspension with respect to p; she has become an *epistemic abstainer* with respect to p, we may say. The fact of Betty’s abstention is relevant to Alf in two ways. First, once Betty suspends judgment, Alf no longer has a disagreement with Betty of the kind that would engage PS with respect to p. So Alf can sustain his belief that p. To be sure, there is still a disagreement between Alf and Betty, but now it is a disagreement concerning whether to suspend judgment regarding p (Betty will present this new disagreement to Alf shortly). Second, note that part of what compels Betty to epistemically abstain is her observation that an epistemic

¹⁴ David Christensen, “Epistemology of Disagreement: The Good News,” *Philosophical Review* 116 (2007): 203.

¹⁵ Earl Conee, “Peerage,” *Episteme* 6 (2009): 315.

¹⁶ Conee, “Peerage,” 322.

peer, Alf, stays adamant in his belief that *p*; but if Betty must take Alf's *immovability* about *p* as evidence about the insufficiency of her own evidence, then Alf may likewise take Betty's *movability* as evidence in favor of the *strength* of his evidence. We imagine Alf reasoning as follows:

If a peer's disagreement is enough to defeat one's reasons, then my peer's movement from full-bore disagreement to suspension of judgment should also be an indicator of the (insufficient) quality my peer's reasons.

In other words, because Alf does not blink, Betty must suspend judgment; and because Betty suspends judgment, Alf may have even more evidence for his view. Once Betty suspends, Alf has one fewer epistemic peer who believes not-*p*, precisely because of her concerns that she had no reason to hold she was not in error. And so, in following PS, Betty moves from being an *epistemic abstainer* to being an *epistemic martyr*, all because of Alf's immovability.

Something has gone awry. It seems that Alf isn't playing fair—Alf improves his epistemic position by simply waiting for Betty to weaken her view. In refusing to apply PS, Alf is being an epistemic free-rider. So let us imagine that Betty calls foul. Betty says, "*Alf, you must suspend judgment with respect to p, too!*" But notice that we now have moved the disagreement from (i) *whether p* to (ii) *what the proper propositional attitude toward p should be*. Alf holds that he is justified in believing that *p*, and Betty holds that Alf should suspend judgment. But, now, Feldman's view would require Betty to apply PS *to this (new) disagreement*. That is, Betty must suspend judgment about whether Alf must suspend judgment with regard to *p*. And so Betty must weaken her claim that Alf is breaking an epistemic rule to a suspension of judgment with respect to Alf's rule following. Additionally, it seems that because Betty and Alf disagree about the application of PS, Betty should suspend judgment about whether she should suspend judgment about *p*.

The discussion between Alf and Betty may then go on like this for several rounds, and potentially forever. At each level, Betty will not have a substantive reason for holding that Alf is unreasonable for sustaining his belief. And she will have reason to suspend judgment about the propriety of her previous abstemious moves. Betty, given Alf's disagreement on each level, will be unable to hold positively that Alf is wrong and she is right on any of the levels. She only can stammer in disbelief, immobilized and in a state of perpetual epistemic *suspension*. Reminding Alf of PS, if Alf is tenacious, only *deepens* Betty's martyrdom.

Meanwhile, things continue to get better for Alf. He is, by his lights at least, *prima facie* justified in his belief that *p*, and further, there are now no peers who dissent. So Betty, following the standing evidence and the social reflections on the

quality of that evidence, should now *come to believe* *p*. That is, Betty, although presently in suspense, should see that Alf believes *p* unopposed. Betty should reason thusly:

Since dissent among peers is a defeater, the absence of peer dissent with respect to p improves the case for p.

So Betty should come to believe that *p*. That is, if peer disagreement is enough to overturn Betty's originally well-thought out reasons, they must, absent her own defeating reasons, be good enough for warrant her assent. Feldman, remember, had argued that peer disagreement is a reason to suspend judgment precisely because the peer's contrary beliefs stand as evidence that the subject's view is false¹⁷ Betty, now that there is no contrary evidence to Alf's view, as she has suspended belief with regard to *p*, now has evidence that not-*p* is true. She should proportion her belief to her evidence. And so Betty rises from her epistemic martyrdom, but now as an *epistemic convert*.

Surely Betty will find this abstention-to-martyrdom-to-conversion experience puzzling. And we do, too. Indeed, it is hard to imagine how such contortions could be a sign of epistemic responsibility (to say nothing of psychological health). The simple fact of Alf's tenacity produces for Betty a sufficient reason for her to adopt his view. Thus Betty's epistemic virtue must succumb to Alf's epistemic vice.

Perhaps then we should rule out free-riders by revising Feldman's view to say that PS applies only in cases in which *both* parties to a disagreement reciprocally suspend judgment. This requirement of reciprocity among epistemic peers may avoid the dizzying shift from martyr to convert; however, it also suggests that PS cannot serve as a response to all cases of disagreement between epistemic peers. In particular, it fails in just those deep disagreements that generate the problem of peer disagreement in the first place. Deep disagreement arises when neither party can agree on the proper attitude to take towards *p* or how to further arbitrate what divides them. In such cases, it will not help if one of the parties suspends judgment with regard to *p*, because this will simply relocate the disagreement: What was once a disagreement about whether *p* becomes a disagreement about whether to suspend judgment with respect to *p*. As we have already seen, one can remain committed to PS in such cases only on pains of becoming an epistemic martyr and ultimately a convert.

¹⁷ Richard Feldman, "Evidentialism," 331.

Epistemic martyrdom occurs in cases in which PS is applied *asymmetrically*, but something similar occurs even when PS is applied *symmetrically*. If *both* sides apply PS when a second-order disagreement arises, all believers will be martyred to suspense, for suspenders dictate the epistemic requirements in the situation for all involved. To see this, let us view the second order disagreement from the perspective of Charles. Charles disagrees with Betty about *p*, just like Alf does. And Charles disagrees with Betty over the propriety of belief or suspension with regard to *p*. However, with this second-order disagreement, he *abides by PS*. So Charles suspends judgment with regard to suspending judgment with regard to *p*. But it seems to Charles that *this is simply to suspend judgment with regard to p*. He reasons that one could not both believe that *p* and suspend judgment with regard to whether to suspend judgment with regard to *p*. According to PS, then, any believer must suspend belief in the face of any peer who suspends judgment. One could avoid this worst-case outcome only by denying that Charles must suspend with regard to *p* when he suspends with regard to suspending about *p*. But this seems inconsistent with overtly holding that *p*: if Charles thinks he is holding that *p* justifiably, he would not suspend judgment about whether he should suspend judgment. That is, he should think it *false* that he should suspend judgment. But since he has suspended judgment about whether he should suspend judgment, he, it seems, has undone his belief.

In short, if everybody plays by Feldman's rules, or those articulated by the broader versions of epistemic abstemiousness, those who suspend judgment with regard to any issue will dictate the epistemic duties of all their peers. Imagine a group of epistemic peers who all believe that *p*. Now introduce to that group an epistemic peer who suspends judgment with regard to *p*. On Feldman's abstemious principles, the entire group must now suspend judgment, regardless of how deeply held the belief is. To Charles, this has the appearance of submitting his deeply held beliefs to the whims of those who suspend their beliefs at the first whiff of disagreement. For better or worse, all is not lost for the believers. Simply add to the group an epistemic peer who ignores this application of PS and stalwartly continues on with her belief that not-*p*. For reasons we provided above, the entire group would now have reason to convert to the stalwart's view. So Feldman's epistemically reasonable believers become martyrs, and the tenacious win easy converts. The trouble with PS, and abstemiousness more broadly, is that, in spite of its broad-minded intentions and overtly anti-dogmatic aim, it recommends dogmatism.

This is indeed a troubling result. PS is intuitively appealing precisely because the alternative of dogmatically holding on to one's belief *in spite of* peer

disagreement seems unreasonable. One need not be an unscrupulous epistemic free-rider in search of easy converts, however, to think that dogmatism is preferable to becoming an epistemic martyr. The requirement that one subject oneself to epistemic martyrdom—or even worse, conversion—simply because of peer disagreement seems far too demanding. Our deeply held beliefs are not the sorts of things that we can simply give up on at a moment's notice. We suspect that part of what belief is to be committed to it in way that prevents one from seeing it as so easily disposable. It is hard then to be committed to PS while maintaining one's *integrity* as a believer. We are concerned that this is the case for abstemious commitments across the board. But our cautionary tales above suggest that Feldman's PS—and perhaps epistemic abstemiousness as such—is deeply at odds with how we view ourselves as cognitive agents.

EPISTEMIC CLOSURE AND SKEPTICISM

John A. BARKER, Fred ADAMS

ABSTRACT: Closure is the epistemological thesis that if S knows that P and knows that P implies Q, then if S infers that Q, S knows that Q. Fred Dretske acknowledges that closure is plausible but contends that it should be rejected because it conflicts with the plausible thesis: Conclusive reasons (CR): S knows that P only if S believes P on the basis of conclusive reasons, i.e., reasons S wouldn't have if it weren't the case that P. Dretske develops an analysis of knowing that centers on CR, and argues that the requirement undermines skepticism by implying the falsity of closure. We develop a Dretske-style analysis of knowing that incorporates CR, and we argue that this analysis not only accords with closure, but also implies it. In addition, we argue that the analysis accounts for the prima facie plausibility of closure-invoking skeptical arguments, and nonetheless implies that they are fallacious. If our arguments turn out to be sound, the acceptability of Dretske's analysis of knowing will be significantly enhanced by the fact that, despite implying closure, it undermines closure-based skepticism.

KEYWORDS: knowledge, closure, conclusive reasons, skepticism, Dretske

1. Introduction

It seems clear that deductively valid inferences from known premises can usually augment knowledge. For example, if S knows that a certain animal, X, is a zebra, and S knows that X's being a zebra implies X's being a mammal, S can acquire inferential knowledge that X is a mammal. Closure is the epistemological thesis that such augmentation of knowledge is always possible:

Closure: If S knows that P and S knows that P implies Q, then if S infers that Q from the premises that P and that P implies Q, S knows that Q.¹

¹ Fred Dretske formulates the thesis in the following way: "Closure is the epistemological thesis that if S knows that P is true and knows that P implies Q, then, evidentially speaking, this is enough for S to know that Q is true." (Fred Dretske, "The Case against Closure," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 13.) In this paper, we focus on versions of closure that at least implicitly

According to Fred Dretske, "... closure sounds like an eminently plausible principle. Everything else being equal, then, we ought to keep it. But everything else *isn't* equal."² He contends that closure conflicts with the following plausible thesis:

Conclusive reasons (CR): S knows that P only if S believes P on the basis of *conclusive reasons*, i.e., reasons S wouldn't have if it weren't the case that P.³

"If knowledge is belief based on the kind of conclusive reasons I describe in Dretske (1971) ..., then closure fails. Things turn out this way because one can have conclusive reasons ... for P ... without having conclusive reasons for known consequences of P."⁴ For example, it seems plausible that one can have conclusive reasons for believing that the animals one sees in the zebra pen of a zoo are zebras without having conclusive reasons for believing "that these animals are not mules cleverly disguised by the zoo authorities to look like zebras."⁵ Dretske develops an

involve competent inference, as these versions seem to be the most plausible ones. For present purposes, closure is to be construed as: If S knows at (time) T that P and S knows at T that P implies Q, then if S competently infers at T that Q immediately from the premises that P and that P implies Q, S knows at T that Q. See, e.g., John Hawthorne, "The Case for Closure," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 29, and Timothy Williamson, *Knowledge and Its Limits* (Oxford: Oxford University Press, 2000), 117, for discussion of the role of competent inference, and see, e.g., Steven Luper, "The Epistemic Closure Principle," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2010 Edition), <http://plato.stanford.edu/archives/sum2010/entries/closure-epistemic/>, for discussion of different versions of closure and for references to relevant literature.

² Dretske, "Closure," 18.

³ For present purposes, CR is to be construed as: S knows that P only if S believes P on the basis of one or more reasons, R, that are such that R wouldn't be the case if it weren't the case that P. The notion of believing something *on the basis of a reason* will be explicated in Section 2. S will be said to *have* a reason, R, for believing P iff R is the case and R is at least one of S's reasons for believing P.

⁴ Dretske, "Closure," 19. For Dretske's views regarding closure and the conclusive reasons requirement, see, e.g., Fred Dretske "Epistemic Operators," *Journal of Philosophy* 67 (1970): 1007–1023, "Conclusive Reasons," *Australasian Journal of Philosophy* 49 (1971): 1–22, "Contrastive Statements," *Philosophical Review* 81 (1972): 411–430, "Closure," and "Reply to Hawthorne," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell, 2005), 43–46.

⁵ Dretske, "Operators," 1016.

analysis of knowing that centers on CR, and argues that the requirement undermines skepticism by implying the falsity of closure.⁶

In view of the plausibility of closure, however, Dretske acknowledges that there are substantial costs associated with rejecting it: “These costs are, I admit, significant. I would not be willing to pay this price if I thought there were alternatives that were less expensive.”⁷ We think there may be a *cost-free* way to defend CR. In this paper, we develop a Dretske-style analysis of knowing that incorporates CR, and we argue that this analysis not only accords with closure, but also implies it. In addition, we argue that the analysis accounts for the *prima facie* plausibility of closure-invoking skeptical arguments, and nonetheless implies that they are fallacious. According to Dretske, rejection of closure is “not just *a* way to avoid skepticism (most philosophers would agree with this) but the *only* way to avoid skepticism.”⁸ If our arguments turn out to be sound, however, the acceptability of Dretske’s analysis of knowing will be significantly enhanced by the fact that, despite implying closure, it undermines closure-based skepticism.⁹

2. A Dretske-style Analysis of Knowing

We agree with Dretske that CR is plausible in its own right, independently of any capacity it may have to discredit skepticism by ruling out closure:

As a historical footnote, I wasn’t led to deny closure because it represented a way around skepticism. I was led to it because it was a result of what I took to be a plausible condition on the evidence (justification, reasons) required for knowledge. If your reasons for believing P are such that you *might* have them

⁶ See Dretske, “Conclusive,” 12 ff., for his analysis of knowing, and see, e.g., Dretske, “Closure,” for his arguments against skepticism. We will use the term ‘skepticism’ to refer to the philosophical view that little or no genuine knowledge exists, and we will be concerned exclusively with closure-based varieties of skepticism.

⁷ Dretske, “Hawthorne,” 43.

⁸ Dretske, “Closure,” 18.

⁹ This paper is a product of an ongoing collaborative effort focused on issues involving knowledge, and neither of us wholeheartedly endorses all of the theses defended herein. (Indeed, some of these theses are not fully in accord with theses we defended in an earlier product of our collaboration, F. Adams, J. Barker, and J. Figurelli, “Towards Closure on Closure,” 2010 (manuscript).) Nevertheless, we think that the arguments we present for these theses constitute worthwhile contributions to current debates about the issues. In this paper, we focus on a Dretske-style analysis of knowing and on a several competing analyses that incorporate CR; in a planned sequel to this paper, we will discuss numerous competing analyses, including many that do not incorporate CR.

when P is false, then they aren't good enough to *know* that P is true. You need something more. That is why you can't know you are going to lose a lottery just because your chances of losing are 99.99 percent. Even with those odds, you still might win (someone with those odds against him *will* win). That is why you can't learn – can't come to know – that P is true if all you have to go on is the word of a person who might lie about whether or not P is so. This is just another way of saying that knowledge requires reasons or evidence (in this case, testimony) you wouldn't have if what you end up believing were false. You can learn things from people, yes, but only from people who wouldn't say it unless it were true.¹⁰

We also think that Dretske's analysis of knowing is plausible in its own right, independently of its import regarding skepticism and closure:

S has conclusive reasons, R, for believing P iff:

(A) R is a conclusive reason for P ...,

(B) S believes, without doubt, reservation, or question, that P is the case and he believes this on the basis of R,

(C) (i) S knows that R is the case or

(ii) R is some experiential state of S (about which it may not make sense to suppose that S *knows* that R is the case; at least it no longer makes much sense to ask *how* he knows).

With only minor embellishments, to be mentioned in a moment, I believe that S's having conclusive reasons for believing P is *both* a necessary and a sufficient condition for his knowing that P is the case. The appearance of the word 'know' in this characterization (in (Ci)) does not render it circular as a characterization of knowledge since it can be eliminated by recursive application of the three conditions until (Cii) is reached.¹¹

This analysis of knowing employs the notion of *believing something on the basis of a reason*, a notion we explicate as follows:

Epistemic-basing: S believes P on the basis of a reason, R, iff: either (i) R is at least one of S's reasons for believing P, and R is an experiential state of S; or (ii)

¹⁰ Dretske, "Hawthorne," 43-44.

¹¹ Dretske, "Conclusive," 12-13. The "minor embellishments" referred to in this passage aren't relevant to present concerns.

S's believing R to be the case is at least one of S's reasons for believing P, and S knows that R is the case.¹²

In light of epistemic-basing, the following analysis of knowing seems plausible:

Dretske-style analysis of knowing (DAK): S knows that P iff S believes P on the basis of conclusive reasons.¹³

DAK accords with Dretske's view that "S's having conclusive reasons for believing P is *both* a necessary and a sufficient condition for his knowing that P is the case."¹⁴

We argue that: (i) DAK specifies correct necessary and sufficient conditions for knowledge itself, which we refer to as *generic* knowledge; (ii) several important *kinds* of knowledge – we call them *species* of knowledge -- can be delineated by supplementing the conditions specified by DAK; (iii) distinct versions of closure hold for generic knowledge and for certain species of knowledge; (iv) closure-invoking skeptical arguments are fallacious because they exploit confusions pertaining to generic knowledge on the one hand, and to various species of knowledge on the other; (v) DAK is superior to several competing analyses, and (vi) as a result, the acceptability of Dretske's analysis of knowing is significantly enhanced.

¹² The notion of *a subject's reasons for believing something* will be employed herein as a primitive notion. We will discuss this notion, and the closely related notion of believing something on the basis of a reason, in the sequel. See, e.g., Marshall Swain, *Reasons and Knowledge* (Ithaca: Cornell University Press, 1981), for an in-depth discussion of these notions, and see, e.g., Keith Allen Korcz, "The Epistemic Basing Relation," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2010 Edition), <http://plato.stanford.edu/archives/sum2010/entries/basing-epistemic/>, for a survey of theories about them. Epistemic-basing is an abbreviation of: S believes P on the basis of a reason, R, iff: either (i) R is at least one of S's reasons for believing P, and R consists of one or more experiential states of S, or (ii) S's believing R to be the case is at least one of S's reasons for believing P, S knows that R is the case, and *S's knowing this doesn't presuppose that P*, or (iii) R consists of a combination of reasons that satisfy the conditions specified by (i) and (ii). The function of the italicized clause in this thesis will be explained in the sequel.

¹³ DAK is to be construed as an abbreviation of: S knows that P iff S believes P on the basis of one or more reasons, R, that are such that R wouldn't be the case if it weren't that case that P. The sentential operator "if it weren't the case that P, it wouldn't be the case that Q," which is employed herein as a primitive operator, will be discussed at length in the sequel.

¹⁴ Dretske, "Conclusive," 13.

The notions of *necessary condition* and *sufficient condition* we employ can be explicated as follows: (i) its being the case that P is a *necessary condition* for its being the case that Q iff it is necessarily the case that if it weren't the case that P, then it wouldn't be the case that Q; and (ii) its being the case that P is a *sufficient condition* for its being the case that Q iff it is necessarily the case that if it weren't the case that Q, it wouldn't be the case that P. We do not presuppose that the fact that something holds of necessity implies either that it holds of *logical* necessity or that its holding is knowable *a priori*.¹⁵ Furthermore, in arguing that DAK specifies correct necessary and sufficient conditions for knowledge, we aren't attempting to challenge views like Timothy Williamson's that "the concept *knows* cannot be analyzed into more basic concepts"¹⁶ and that "the pursuit of analyses is a degenerating research programme."¹⁷ While we will follow the standard practice of referring to DAK and its competitors as *analyses of knowing*, we will take no stand in this paper on the highly controversial issues associated with the nature and prospects of so-called *conceptual analysis*.¹⁸

The principal theses we defend—that DAK implies closure and that it nonetheless undermines closure-based skepticism—seemed implausible to us in the initial stages of our investigation, and we anticipate that these theses will, at first glance, seem implausible to many of our readers. Not surprisingly, we experienced considerable difficulty in building a strong case for acceptance of the theses, and found ourselves agreeing with John Hawthorne, one of Dretske's many critics, that "if there were some easily accessible locus of reflective equilibrium in the vicinity, we would surely have reached it by now."¹⁹ Accordingly, we have

¹⁵ We do not claim that the notions of *necessary condition* and *sufficient condition* we employ are the ones that are most commonly employed by other theorists. See, e.g., Andrew Brennan, "Necessary and Sufficient Conditions," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Fall 2008 Edition), <http://plato.stanford.edu/archives/fall2008/entries/necessary-sufficient/>, for discussion of various conceptions of necessary conditionship and sufficient conditionship..

¹⁶ Williamson, *Knowledge*, 33.

¹⁷ Williamson, *Knowledge*, 31.

¹⁸ See, e.g., Robert Shope, *The Analysis of Knowing: A Decade of Research* (Princeton: Princeton University Press, 1983), 34 ff. and Michael Beaney, "Analysis," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2009 Edition), <http://plato.stanford.edu/archives/sum2009/entries/analysis/>, for discussions of these issues, which we will address in the sequel. See also John Hyman "Knowledge and Evidence," *Mind* 115 (2006): 891-916, for arguments against Williamson's view and in favor of an analysis of knowing that, as we will show in the sequel, resembles DAK.

¹⁹ Hawthorne, "Closure," 27.

adopted the strategy of presenting our case in a highly methodical fashion characterized by precise definitions of key terms and by argumentation that is largely formal. While we think that our arguments succeed in establishing that DAK implies closure and nonetheless undermines closure-based skepticism, we are considerably less confident that they establish DAK's superiority to its competitors. Indeed, we will end this paper with an admission that some of these competitors are, in our opinion, still in the running.

3. Generic Knowledge and Species of Knowledge

The following cases illustrate ways in which distinguishing between generic knowledge and various species of knowledge lends support to DAK and to closure, and yet serves to undermine closure-based skepticism. Jimmy and his mother, Lisa, see a certain aquatic animal, X, swimming in the ocean. Jimmy believes X to be a porpoise because it appears to him to be one; in other words, his having this experience is at least one of his reasons for believing that X is a porpoise.²⁰ Epistemic-basing implies that his belief is based on his having the experience. If X weren't a porpoise, it would be a seal, a shark, or something else that wouldn't appear to him to be a porpoise. Consequently, his having the experience is a conclusive reason for his belief, and DAK implies that the belief qualifies as generic knowledge. Now Jimmy has never seen a dolphinfish, and if X were a dolphinfish, it might appear to him to be a porpoise. Nevertheless, there are no dolphinfish in the vicinity, and if X weren't a porpoise, it wouldn't be such a fish. Knowing that X is a porpoise, and knowing that X's being a porpoise implies its not being a dolphinfish, Jimmy infers that X isn't such a fish. Does he thereby acquire generic knowledge that X isn't a dolphinfish?

Consider the plausible thesis:

Inferential-reasons: If S infers that Q from a premise that P, then at least one of S's reasons for believing Q is S's believing P.

Assuming that this thesis is true, it follows that at least one of Jimmy's reasons for believing that X isn't a dolphinfish is his believing X to be a porpoise. Since he knows X is a porpoise, epistemic-basing implies that he believes X isn't a

²⁰ The term 'because' will (almost invariably) be used in this paper to refer to a subject's *reason* or *reasons* for believing something, for intending, doubting, or wanting something, or for performing some (intentional) action. This use of the term will be discussed in more detail in the sequel. Expressions of the form 'x appears to S to be F' and 'it appears to S that x is F' will be used as equivalents.

dolphinfish on the basis of a reason consisting of X's being a porpoise. X wouldn't be a porpoise were it a dolphinfish. Consequently, X's being a porpoise is a conclusive reason for his conclusion belief, and DAK implies that the belief qualifies as generic knowledge. Thus, with the help of epistemic-basing and inferential-reasons, DAK implies closure.^{21 22}

Lisa, who is an ichthyologist, believes that X is a porpoise *rather than* a dolphinfish because X appears to her to be a porpoise. Epistemic-basing implies that her belief is based on her having this experience. X wouldn't appear to her to be a porpoise were it not one, and X wouldn't appear to her to be a porpoise were it a dolphinfish. Consequently, her having the experience is a conclusive reason not only for her believing X is a porpoise, but also for her believing X isn't a dolphinfish, and DAK implies that both of these beliefs qualify as generic knowledge.²³ Lisa's belief that X is a porpoise rather than a dolphinfish also qualifies as knowledge, for X's appearing to her to be a porpoise functions as a *differentiator*, i.e., a reason that enables her to distinguish between the competing possibilities, X's being a porpoise and X's being a dolphinfish. She possesses what we'll call *contrastive* knowledge, which we explicate as follows:

Contrastive-knowing: S knows that x is A rather than B iff x's being A entails x's not being B, and S believes that x is A and not B on the basis of a *contrastively*

²¹ The derivation of closure from DAK, epistemic-basing, and inferential-reasons will be discussed in detail in the sequel. The derivation involves the plausible presupposition that if it's the case both that P and that P implies Q, then it wouldn't be the case that P if it weren't the case that Q. In virtue of this presupposition, a stronger form of closure can be derived: If S knows that P, then if P implies Q and S infers that Q from the premises that P and that P implies Q, S knows that Q. We will discuss the role this principle plays in accounting for acquisition of knowledge via non-deductive reasoning.

²² Steven Luper has advanced a similar argument against Dretske's denial of closure: "We might insist that *p* itself is a conclusive reason for believing *q* when we know *p* and *p* entails *q*. After all, assuming *p* entails *q*, if *q* were false so would *p* be. On this strategy we have a further argument for [closure]: if S knows *p* (relying on some conclusive reason R), and S believes *q* because S knows *p* entails *q*, S has a conclusive reason for believing *q*, namely *p* (rather than R), and hence S knows *q*." (Luper, "Epistemic Closure.") Peter Klein has advanced a somewhat similar argument, although he is concerned with knowledge construed as a form of justifiable or defensible belief rather than as a form of belief based on conclusive reasons. (Peter Klein, "Skepticism," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Summer 2010 Edition), <http://plato.stanford.edu/archives/sum2010/entries/skepticism/>.) We argue later in the paper that closure does not hold if knowledge is construed in this way.

²³ We take it for granted that her believing that X is a porpoise rather than a dolphinfish involves her believing X is a porpoise and her believing X isn't a dolphinfish.

conclusive reason, R, i.e., (i) R wouldn't be the case if it weren't the case that x is A, (ii) R wouldn't be the case if it were the case that x is B, and (iii) R doesn't entail either x's being A or x's not being B.²⁴

Since X's appearing to Lisa to be a porpoise constitutes a contrastively conclusive reason for her belief that X is a porpoise rather than a dolphinsfish, contrastive-knowing implies that the belief qualifies as knowledge.

Jimmy is unfamiliar with dolphinsfish, and doesn't know what such animals look like. Even though X wouldn't appear to him to be a porpoise were it not a porpoise, X might appear to him to be a porpoise were it a dolphinsfish. Consequently, he is in no position to acquire knowledge that X is a porpoise *rather than* a dolphinsfish. Knowing that X is a porpoise and knowing that it isn't a dolphinsfish, he infers that X is a porpoise rather than a dolphinsfish. His conclusion belief is not based on a reason that can qualify as a differentiator enabling him to distinguish between the competing possibilities, X's being a porpoise and X's being a dolphinsfish. X's appearing to him to be a porpoise cannot qualify, for he might have this experience were X a dolphinsfish. X's being a porpoise cannot qualify, for it entails X's not being a dolphinsfish. Similarly, the conjunctive state of affairs consisting of X's both appearing to be a porpoise and being a porpoise cannot qualify.²⁵ Hence, he lacks a contrastively conclusive reason for his conclusion belief, and contrastive-knowing implies that he doesn't know that X is a porpoise rather than a dolphinsfish. It seems plausible that only the following restricted version of closure holds for this species of knowledge:

Contrastive-knowledge-closure: If (i) S knows that x is A; (ii) S knows that x's being A entails x's not being B; (iii) S infers that x is A rather than B; and (iv) S believes x is A on the basis of a contrastively conclusive reason for believing that x is A and not B; then S has contrastive knowledge that x is A rather than B.

²⁴ This thesis is to be construed as an abbreviation of: S knows that it's the case that P rather than the case that Q iff: P entails not-Q, and S believes that P and not-Q on the basis of a *contrastively* conclusive reason, R, i.e., (i) R wouldn't be the case if it weren't the case that P, (ii) R wouldn't be the case if it were the case that Q, and (iii) R doesn't entail either P or not-Q. The function of Clause (iii) will be clarified in the next paragraph.

²⁵ In ensuring that these states of affairs cannot qualify as differentiators, Clause (iii) of contrastive-knowing helps account for the fact that Lisa's knowing that X is a porpoise *rather than* a dolphinsfish constitutes an epistemic achievement that is superior to Jimmy's knowing that X is a porpoise *and not* a dolphinsfish, i.e., his knowing that X is a porpoise and X is not a dolphinsfish. (We don't wish to suggest that the expression 'and not' is never used in the sense of 'rather than'.)

Besides possessing contrastive knowledge that X is a porpoise rather than a dolphinfish, Lisa possesses what we'll call *experiential* knowledge that X isn't a dolphinfish – her knowledge-qualifying belief is based on a conclusive reason, X's appearing to her to be a porpoise, which consists of an experiential state.²⁶ Jimmy lacks such knowledge – X's being a porpoise, which is the only conclusive reason he has for believing that X isn't a dolphinfish, doesn't consist of an experiential state. Lisa also possesses what we'll call *defensible* knowledge that X isn't a dolphinfish – her knowledge-qualifying belief is based on a conclusive reason that makes the belief *defensible*, i.e., justifiable from her own perspective.²⁷ Jimmy lacks such knowledge – owing to his lack of familiarity with dolphinfish, his belief that X isn't a dolphinfish isn't defensible. It seems plausible that only the following restricted versions of closure hold for these species of knowledge:

Experiential-knowledge-closure: If (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes P on the basis of an experiential reason that is a conclusive reason for believing Q; then S has experiential knowledge that Q.

Defensible-knowledge-closure: If (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes that P on the basis of a reason that makes believing that Q defensible; then S has defensible knowledge that Q.

Some of DAK's competitors supplement CR with additional requirements that have the effect of equating generic knowledge with contrastive knowledge, experiential knowledge, or defensible knowledge. For example, the following CR-incorporating analyses of generic knowledge are competitors of DAK:

CR+contrastivity: S knows that P iff S believes P on the basis of a conclusive reason, R, that is such that, for any incompatible Q that might be the case were P

²⁶ In delineating various *kinds* (or *species*) of knowledge, we are making no claims about the existence of context-dependent variations in *senses* of the term 'know' or in *standards* for the term's applicability. Hence, we are not advocating adoption of *epistemic contextualism*. See, e.g., Patrick Rysiew, "Epistemic Contextualism," *The Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta (Spring 2009 Edition), <http://plato.stanford.edu/archives/spr2009/entries/contextualism-epistemology/>, for discussion of epistemic contextualism and for references to relevant literature.

²⁷ For present purposes, the notion of defensible belief, which will be discussed in detail in the sequel, can be explicated as follows: S *defensibly* believes P on the basis of R iff S's believing P on the basis of R would be more reasonable from S's own perspective than not doing this, were S concerned at the relevant time only with acquiring the truth regarding whether or not P by doing it.

not the case, R can qualify as a contrastively conclusive reason for believing P and not-Q.²⁸

CR+experientiality: S knows that P iff S believes P on the basis of a conclusive experiential reason.

CR+defensibility: S knows that P iff S believes P on the basis of a conclusive reason that makes the belief defensible.

We contend that acceptance of any of these analyses necessitates rejection of closure. As the following considerations suggest, theorists who accept one of these analyses and also accept closure are committed to accepting skepticism. Knowing that X's being a porpoise implies X's not being a dolphinsfish that appears to her to be a porpoise, Lisa infers that X isn't such a fish. Consider the following skeptical arguments:

(A1) If Lisa knows X is a porpoise, then, in virtue of closure, she knows it isn't a dolphinsfish that appears to her to be a porpoise. But she doesn't know X isn't such a fish, for X's appearing to her to be a porpoise doesn't enable her to distinguish between the competing possibilities, X's being a porpoise and its being a dolphinsfish that appears to her to be a porpoise. Consequently, she doesn't know X is a porpoise.

(A2) If Lisa knows X is a porpoise, then, in virtue of closure, she knows it isn't a dolphinsfish that appears to her to be a porpoise. But she doesn't know X isn't such a fish, for X would appear to her to be a porpoise were it a dolphinsfish that appears to her to be a porpoise. Consequently, she doesn't know X is a porpoise.

(A3) If Lisa knows X is a porpoise, then, in virtue of closure, she knows it isn't a dolphinsfish that appears to her to be a porpoise. But she doesn't know X isn't such a fish, for X's appearing to her to be a dolphinsfish doesn't make her conclusion belief defensible. Consequently, she doesn't know X is a porpoise.

Since similar skeptical arguments could be devised to discredit practically any case of apparent knowledge, theorists who adopt one of the above-mentioned analyses of generic knowledge and accept closure appear to be committed to accepting

²⁸ CR+contrastivity implies that Jimmy doesn't know that X is not a dolphinsfish on the basis of X's being a porpoise because there is an incompatible Q, viz. X is a dolphinsfish, which might be (indeed, would be) the case were X a dolphinsfish, and X's being a porpoise, in entailing not-Q, cannot qualify as a contrastively conclusive reason for believing that X isn't a dolphinsfish and not-Q. CR+contrastivity, a version of contrastivism about knowledge that was inspired by Dretske's views about contrastive aspects of knowledge attributions, will be discussed in the sequel and compared with other versions of contrastivism about knowledge.

skepticism. Consequently, the *prima facie* implausibility of skepticism constitutes weighty evidence against their views.

We suggest that would-be skeptics who adopt such CR-incorporating analyses of generic knowledge will be unsuccessful in building a strong case for acceptance of closure. Given epistemic-basing and inferential-reasons, DAK implies closure, but CR+contrastivity, CR+experientiality, and CR+defensibility imply only the following weaker theses:

CR+contrastivity-closure: If (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes P on a basis that constitutes a contrastively conclusive reason for believing P and Q; then S knows that Q.

CR+experientiality-closure: If (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes P on the basis of an experiential reason that is a conclusive reason for believing Q; then S knows that Q.

CR+defensibility-closure: If (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes P on the basis of a reason that makes believing Q defensible; then S knows that Q.

These theses, which are too weak to support arguments like (A1), (A2) or (A3), are nonetheless strong enough to accommodate the intuitions that the skeptics would depend upon for purposes of building a case for acceptance of closure.²⁹

After exploring the implications of DAK for questions about the structure of a subject's fund of knowledge, we will attempt to show that DAK accounts for the *prima facie* plausibility of closure-invoking skeptical arguments and nonetheless implies that they are fallacious.

4. The Structure of Knowledge

DAK and epistemic-basing suggest that a subject's fund of knowledge can be usefully portrayed as having an edifice-like structure. The fund's foundation is made up of knowledge-qualifying beliefs based immediately on knowledge-

²⁹ An epistemic contextualist might contend that: (1) depending on the intentions, presuppositions, etc. of the speaker, a sentence such as "Lisa knows that X is a porpoise" can express a proposition that possesses truth conditions specified by DAK, by CR+contrastivity, by CR+experientiality, or by CR+defensibility, and (2) the skeptical arguments (A1), (A2) and (A3) are intuitively powerful because they are sound, provided that the knowledge-sentences they contain have the truth conditions specified by CR+contrastivity, by CR+experientiality, or by CR+defensibility, respectively. In the sequel we'll argue that such a contention is untenable because the skeptical arguments presuppose versions of closure that are unacceptable.

sufficing reasons that consist of experiential states, and its superstructure is made up of knowledge-qualifying beliefs based immediately on knowledge-sufficing reasons that consist of known facts, and based mediately on reasons that consist of experiential states.

The following case can be used to illustrate this foundationalist portrayal. At least one of Mia's reasons for believing that what she sees, X, is a bear is its appearing to her that X is a bear. Epistemic-basing implies that she believes X to be a bear on the basis of having this experience. She wouldn't have the experience if X weren't a bear.³⁰ Consequently, her having the experience is a conclusive reason for her believing that X is a bear, and DAK implies that the belief qualifies as knowledge. Knowing that X is a bear and knowing that its being a bear implies its being an animal, she infers that X is an animal. It's plausible that she knows that X is an animal, and her knowing this is implied by closure.

Her knowing that X is an animal is also implied by DAK, epistemic-basing, and inferential-reasons. Inferential-reasons implies that Mia's believing that X is a bear is at least one of her reasons for believing that X is an animal. Since she knows X is a bear, epistemic-basing implies that she believes X to be an animal on the basis of its being a bear. X wouldn't be a bear were it not an animal. Consequently, X's being a bear is a conclusive reason for her believing X to be an animal, and DAK implies that she knows that X is an animal. Thus, with the help of epistemic-basing and inferential-reasons, DAK implies closure.

Mia's knowledge-qualifying belief that X is a bear is part of the foundation of her fund of knowledge, for this belief is based immediately on a knowledge-sufficing reason, X's appearing to her to be a bear, that consists of an experiential state. Her knowledge-qualifying belief that X is an animal is part of the superstructure of her fund of knowledge, as this belief is based immediately on a knowledge-sufficing reason, X's being a bear, that she knows to be the case, and is based mediately on a reason, X's appearing to her to be a bear, that consists of an experiential state.³¹

³⁰ In saying this, we mean that if it weren't the case that what she sees is a bear, it wouldn't be the case that what she sees appears to her to be a bear. What she actually sees is a bear, and presumably it's impossible for the bear she sees to be anything other than a bear. But it's possible that she sees a non-bear instead of a bear; accordingly, it's possible that what she sees is a non-bear that doesn't appear to her to be a bear.

³¹ A reason, R, is an *immediate reason* of S's for believing P, let's say, iff there is no other reason, R', of S's for believing P that is such that R is a reason of S's for possessing R', and R' is a reason of S's for believing P. A state of S is *directly accessible* to S, let's say, iff it can in principle be S's immediate reason for believing that S is in the state. A state of S is an *experiential state* of S, let's

Although this foundationalist portrayal of Mia's fund of knowledge can be helpful, it can be misleading for the following reasons: (i) The portrayal may suggest that in order to acquire a fund of knowledge, Mia must become aware of her experiential states and beliefs, and endeavor to ground her foundational-level beliefs upon these states and to construct a superstructure of additional beliefs that are supported by these beliefs and states. If Mia is a young child or an autistic person who lacks the ability to become aware of her experiential states and beliefs and to engage in such endeavors, the portrayal may suggest that her beliefs about X don't qualify as knowledge. And if her becoming aware of her experiential states and beliefs involves her acquiring knowledge about them, the portrayal may suggest that acquisition of knowledge requires prior possession of knowledge. Epistemic-basing and DAK imply that Mia's beliefs about X qualify as knowledge even if she lacks the ability to become aware of her experiential states and beliefs and to engage in such endeavors.³² (ii) The portrayal may suggest that Mia's belief that X is an animal isn't 'sufficiently supported,' and therefore doesn't qualify as knowledge, unless X's appearing to her to be a bear constitutes a knowledge-sufficing reason for it. Accordingly, the portrayal may make the following thesis seem acceptable:

say, iff it's a directly accessible state of S that has the natural function of inducing belief. For example, its sensorially seeming to S that P, its intellectually seeming to S that P, etc. are experiential states of S, for they are directly accessible states of S that have the natural function of inducing belief. In contrast, S's hoping that P, S's fearing that P, S's desiring that P, S's feeling sad that P, etc. are directly accessible states of S that have natural functions other than that of inducing belief. (The question of whether all directly accessible states of S are mental states of S will be discussed in the sequel.)

³² S is *directly aware* of its being the case that P, let's say, iff its being the case that P is an immediate reason of S's for believing P. If its being the case that P is an immediate reason of S's for believing P, it *ipso facto* constitutes a conclusive reason of S's for believing P, for if it weren't the case that P, S wouldn't have this reason for believing P. Hence, DAK implies that if S is directly aware of its being the case that P, S knows that P. If a state of S is directly accessible to S, then it's possible in principle for S to become directly aware of being in the state. Since X's appearing to Mia to be a bear is an experiential state, it's directly accessible to her, and therefore she can in principle become directly aware of being in this state. (She may, of course, fail to do so because she lacks the requisite concepts, etc.) Thus, epistemic-basing and DAK can accommodate her becoming aware of X's appearing to her to be a bear. Nevertheless, these theses imply that her doing so isn't necessary for her acquiring knowledge that X is a bear and that X is an animal; consequently, the theses don't give rise to regress problems.

Epistemic-transitivity: If R is a knowledge-sufficing reason of S's for believing P, and its being the case that P is a knowledge-sufficing reason of S's for believing Q, then R is a knowledge-sufficing reason of S's for believing Q.

Acceptance of this thesis, however, is incompatible with acceptance of DAK and epistemic-basing. The bear Mia sees is in the bear enclosure of a zoo. Whenever the zookeepers remove the bear from the enclosure, they replace it with an ostrich. Accordingly, if what Mia sees weren't a bear, it would be an animal that wouldn't appear to her to be a bear. Now if it so happened that all of the zoo's ostriches died, the zookeepers would place an animated bear replica in the enclosure. Consequently, if what Mia sees weren't an animal, it would be a bear replica that might appear to her to be a bear. Since what Mia actually sees, X, is a bear that wouldn't appear to her to be a bear were it not one, DAK and epistemic-basing imply that X's appearing to her to be a bear is a knowledge-sufficing reason for her believing that X is a bear, and that X's being a bear is a knowledge-sufficing reason for her believing that X is an animal, even though X's appearing to her to be a bear isn't a knowledge-sufficing reason for her believing that X is an animal. Thus, DAK and epistemic-basing imply that the epistemic-transitivity thesis is false.

DAK and epistemic-basing suggest that a subject's fund of knowledge can be more appropriately portrayed as a collection of *informational networks*, i.e., networks that contain interconnected beliefs, experiential states, and information-conveying reasons for beliefs. In virtue of being a conclusive reason for Mia's believing that X is a bear, X's appearing to her to be a bear conveys the information that X is a bear, and enables her believing that X is a bear to result in her believing that X is an animal on the basis of this information. This basis constitutes a conclusive reason for the belief, and therefore conveys the information that X is an animal and ensures that the belief qualifies as knowledge. X might be a bear replica that appears to her to be a bear were it not an animal. Hence, X's appearing to her to be a bear, which is a mediate reason for her believing that X is an animal, isn't a conclusive reason for her believing this, and therefore doesn't convey the information that X is an animal.³³ While the

³³ As we are using the expression 'conveys the information', a reason, R, of S's for believing P *conveys the information* that P iff R wouldn't be the case if it weren't the case that P. Although a reason's *conveying information* is closely related to its *carrying information* (as the latter property is defined by Dretske), the two properties may not be identical. See, e.g., Fred Dretske, *Knowledge and the Flow of Information* (Cambridge: MIT Press, 1981), for a discussion of the notion of *carrying information*.

foundationalist portrayal of Mia's fund of knowledge may suggest that this mediate reason doesn't 'sufficiently support' the belief that X is an animal, the informationalist portrayal suggests that there is simply no need for this mediate reason to convey the information that X is an animal -- X's being a bear, which is an immediate reason for her belief that X is an animal, is a conclusive reason for this belief, and therefore conveys the information that X is an animal and ensures that the belief qualifies as knowledge.³⁴

5. Skepticism and Epistemic-Transitivity

In making the epistemic-transitivity thesis seem acceptable, the foundationalist portrayal of a subject's fund of knowledge may create the impression that the requirements specified by DAK are insufficient for knowledge, and should be augmented along the following lines. S *unassailably* believes P, let's say, iff S believes P on a basis that constitutes a conclusive reason for believing any Q implied by P. Consider the following analysis of knowing:

CR+unassailability: S knows that P iff S unassailably believes P on the basis of a conclusive reason.³⁵

CR+unassailability implies that Mia's superstructure-level belief that X is an animal qualifies as knowledge only if X's appearing to her to be a bear constitutes a conclusive reason for it, for she cannot unassailably believe that X is a bear on the basis of having this experience unless her having it constitutes a conclusive reason not only for X's being a bear, but also for X's being an animal and for everything else implied by X's being a bear. Thus, CR+unassailability implies the epistemic-transitivity thesis. Accordingly, acceptance of CR+unassailability involves acceptance of a view we'll call *strong experiential foundationalism*: beliefs qualify as superstructure-level knowledge only if they potentially qualify as

³⁴ The two portrayals aren't incompatible, and a view that involves acceptance of epistemic-basing and DAK can be classified both as a version of foundationalism and as a version of what can be called *informationalism*: the view that conveyance of information should be accorded a central role in epistemology. The superiority of the informationalist portrayal derives from the fact that it lends no support to the epistemic-transitivity thesis. (The foundationalist and informationalist portrayals concern the structure of a subject's fund of *knowledge* rather than the structure of a subject's fund of *justifiable beliefs*.)

³⁵ The conclusive-reason specification in CR+unassailability is implied by the unassailable-belief specification. Accordingly, CR+unassailability can be expressed as: S knows that P iff S unassailably believes P.

foundational-level knowledge. CR+unassailability implies that in order to qualify as knowledge, Mia's superstructure-level belief that X is an animal must be capable of qualifying as foundational-level knowledge, for it must be the case that if she based her belief that X is an animal directly upon its appearing to her to be a bear, the belief would qualify as foundational-level knowledge. Acceptance of DAK involves acceptance of a weaker view that we'll call *moderate experiential foundationalism*: Beliefs qualify as superstructure-level knowledge only if they are based on facts that are known on the basis of conclusive experiential reasons.

CR+unassailability suffers from drawbacks that don't affect DAK: (i) CR+unassailability has the implausible consequence that Mia doesn't know that X is a bear – her belief that X is a bear doesn't qualify as unassailable, since it's basis, X's appearing to her to be a bear, isn't a conclusive reason for her believing that X is an animal. (ii) CR+unassailability has the implausible consequence that Mia doesn't know that X is an animal – her belief that X is a bear doesn't qualify as knowledge, and therefore her belief that X is an animal isn't based on X's being a bear. (iii) X's being a bear implies X's not being a *bear-doppelganger*, i.e., a non-bear that appears to her to be a bear because she is a victim of massive deception by an evil genius. If CR+unassailability is true, she cannot know that X is a bear on the basis of X's appearing to her to be a bear unless she wouldn't have this experience if X were a bear-doppelganger, for CR+unassailability specifies that her having the experience must be a conclusive reason not only for X's being a bear, but also for everything that X's being a bear implies. Since this condition is obviously unsatisfiable, the requirements for knowledge specified by CR+unassailability are implausibly stringent. (iv) Intuitions supporting CR+unassailability can be accommodated by distinguishing between requirements for knowledge itself, i.e., generic knowledge, and requirements for an important species of knowledge that can be called *unassailable knowledge*: S has *unassailable knowledge* that P iff S unassailablely believes that P on the basis of a conclusive reason. As Cartesians would be apt to contend, it seems plausible that Mia could acquire unassailable knowledge of at least some propositions, e.g., that X appears to her to be a bear, that she thinks X is a bear, that she exists, etc.³⁶ It seems clear,

³⁶ S *thinks* that P, let's say, iff S occurrently believes that P. It's plausible that: (i) Mia can become directly aware of X's appearing to her to be a bear and of her thinking that X is a bear; and (ii) even though X's appearing to her to be a bear implies that it's not the case that X doesn't appear to her to be a bear because she is a victim of massive deception by an evil genius, and even though her thinking that X is a bear implies it's not being the case that she doesn't think that X is a bear because she is a victim of massive deception by an evil genius, she possesses

however, that she cannot acquire unassailable knowledge that X is a bear on the basis of its appearing to her to be one.

A skeptic who accepts CR+unassailability could respond along the following lines:

Assume for the sake of the argument that Mia acquires knowledge that X is a bear on the basis of the conclusive reason consisting of X's appearing to her to be a bear. Knowing that X is a bear and that its being a bear implies its not being a bear-doppelganger, she infers that X isn't a bear-doppelganger. Closure implies that her conclusion belief qualifies as knowledge. But it's implausible that this belief does qualify. While DAK implies that the belief qualifies, CR+unassailability implies that it doesn't. Clearly, then, the requirements specified by DAK are insufficient for knowledge. Although CR+unassailability has the implausible consequences that she doesn't know that X is a bear and that she doesn't know that X is an animal, it nonetheless constitutes an analysis of knowing that is superior to DAK.

It does seem implausible that Mia's conclusion belief that X isn't a bear-doppelganger qualifies as knowledge. But what accounts for this intuition? The skeptic's argument presupposes that the implausibility derives from the belief's not qualifying as unassailable. An anti-skeptic who accepts DAK could proffer the following rebuttal:

It seems clear that Mia's belief that X isn't a bear-doppelganger isn't *defensible*, i.e., justifiable from her own perspective. The very content of her belief -- that X isn't a non-bear that appears to her to be a bear because she is a victim of massive deception by an evil genius -- should, it seems, make it obvious to her that basing the belief upon X's appearing to her to be a bear isn't a reasonable way for her to acquire the truth about the matter. Accordingly, it's likely that the implausibility derives from the belief's not being defensible rather than from its not being unassailable, and this assessment is confirmed by the following considerations: (i) Assume that Mia mistakenly but rationally believes that she is a victim of massive deception by an evil genius, and assume that she mistakenly but rationally believes that an omnipotent deity constantly intervenes on her behalf to ensure that her visual-experience-based beliefs are invariably true. It now seems plausible that her conclusion belief is defensible -- from her own perspective, the belief may even be just as justifiable as her beliefs that X is a bear and that X is an animal. And it now seems plausible that she not only knows that X is a bear and that it's an animal, but also that it isn't a bear-doppelganger. (ii)

unassailable knowledge that X appears to her to be a bear and that she thinks that X is a bear. (Perhaps she can even acquire unassailable knowledge that she exists.)

Assume instead that Mia doesn't have the above-mentioned beliefs about an evil genius and an omnipotent deity, and assume that she is a victim of massive deception by an evil genius, but an omnipotent deity constantly intervenes on her behalf to ensure that her visual-experience-based beliefs are invariably true. It once again seems clear that her belief that X isn't a bear-doppelgänger is indefensible, and it seems implausible that the belief qualifies as knowledge. But the belief qualifies as unassailable, and CR+unassailability, like DAK, implies that she not only knows that X is a bear and that it's an animal, but also that it's not a bear-doppelgänger.

Since CR+unassailability and DAK both imply that defensibility isn't a requirement for knowledge, theorists who accept either of these analyses must attempt to show that intuitions associated with defensibility really aren't germane to judgments regarding the presence or absence of generic knowledge. As we have suggested earlier in the paper, such intuitions can be accommodated by distinguishing between requirements for generic knowledge and requirements for an important species of knowledge we called *defensible knowledge*. S has *defensible knowledge* that P iff S defensibly believes that P on the basis of a conclusive reason.

As the following considerations show, another foundationalism-inspired way of augmenting DAK's conditions—CR+experientiality, i.e., S knows that P iff S believes P on the basis of a conclusive experiential reason—yields a competing analysis of knowing that can undermine the above skeptical argument.³⁷ CR+experientiality, epistemic-basing, and inferential-reasons imply CR+experientiality-closure, i.e., if (i) S knows that P; (ii) S knows that P implies Q; (iii) S infers that Q; and (iv) S believes P on the basis of an experiential reason that is a conclusive reason for believing Q; then S knows that Q.³⁸

An anti-skeptic who accepts CR+experientiality could respond to the above skeptical argument along the following lines:

³⁷ In a personal communication, Dretske suggested that we explore competing analyses like CR+experientiality. CR+experientiality doesn't imply that a knowledge-qualifying belief that P must be based *immediately* upon a conclusive experiential reason, but does imply that any sequence of conclusive reasons leading to a knowledge-qualifying belief that P must contain a conclusive experiential reason for believing P.

³⁸ As noted above, Mia knows that X's being a bear implies its being an animal. Suppose that: (i) she knows this only if she knows it *a priori*, i.e., only if she knows it independently of sensory experience, and (ii) she believes it on the basis of a conclusive reason consisting of its *intellectually* seeming to her that it's so. CR+experientiality implies that her belief qualifies as knowledge. Since it's plausible that both of these suppositions can be true, CR+experientiality seems capable of accommodating her knowing that X's being a bear implies its being an animal.

Assume for the sake of the argument that Mia acquires knowledge that X is a bear on the basis of the conclusive reason consisting of X's appearing to her to be a bear. Knowing that X is a bear and that its being a bear implies its not being a bear-doppelganger, she infers that X isn't a bear-doppelganger. Although closure implies that her conclusion belief qualifies as knowledge, CR+experientiality-closure doesn't have this implausible consequence – even though epistemic-basing and inferential-reasons imply that the belief is based on a conclusive reason consisting of X's being a bear, this reason isn't an experiential state. While DAK has the plausible consequence that her premise belief qualifies as knowledge, it implies closure and therefore has the implausible consequence that her conclusion belief also qualifies. And while CR+unassailability has the plausible consequence that her conclusion belief doesn't qualify as knowledge, it implies closure and therefore has the implausible consequence that her premise belief doesn't qualify. In contrast, CR+experientiality has the plausible consequence that her premise belief qualifies as knowledge and has the plausible consequence that her conclusion belief doesn't qualify. Clearly, then, the requirements specified by DAK are insufficient for knowledge, and those specified by CR+unassailability are so stringent that they lead to skepticism. Thus, CR+experientiality constitutes an analysis of knowing that is superior to these competitors.

CR+experientiality, like CR+unassailability, implies that Mia's superstructure-level belief that X is an animal qualifies as knowledge only if X's appearing to her to be a bear constitutes a conclusive reason for it. Thus, acceptance of CR+experientiality involves acceptance of the epistemic-transitivity thesis and acceptance of strong experiential foundationalism. CR+experientiality implies that in order to qualify as knowledge, Mia's superstructure-level belief that X is an animal must be capable of qualifying as foundational-level knowledge—it must be the case that if she based her belief that X is an animal directly upon its appearing to her to be a bear, the belief would qualify as foundational-level knowledge.

CR+experientiality suffers from drawbacks that don't affect DAK: (i) Since X might be a bear replica that appears to Mia to be a bear were it not an animal, CR+experientiality has the implausible consequence that she doesn't know that X is an animal even though she knows that X is a bear, knows that its being a bear implies its being an animal, and infers that it's an animal. (ii) CR+experientiality is incompatible with a highly plausible principle, viz. closure. (iii) Although it's implausible that Mia's belief that X isn't a bear-doppelganger qualifies as knowledge, the source of this implausibility seems to be the belief's being indefensible rather than the belief's lacking a conclusive experiential basis. Her belief that X is an animal seems to be defensible, and it's plausible that it qualifies

as knowledge even though it lacks a conclusive experiential basis. Assume that Mia's belief that X is a bear is indefensible because she mistakenly but rationally believes that she is a victim of massive deception by an evil genius, but lacks any beliefs that make it reasonable for her to rely on her visual experiences to acquire the truth about the matter. It seems plausible that her belief that X is a bear doesn't qualify as knowledge, despite being based on a conclusive experiential reason. (iv) CR+experientiality implies the epistemic-transitivity thesis, which is closely associated with aspects of the foundationalist portrayal of a subject's fund of knowledge that, from the perspective afforded by the informationalist portrayal, appear to be arbitrary and misleading. In virtue of being a conclusive reason for Mia's believing that X is a bear, X's appearing to her to be a bear conveys the information that X is a bear, and enables her believing that X is a bear to result in her believing that X is an animal on the basis of this information. This basis constitutes a conclusive reason for the belief, and therefore conveys the information that X is an animal and ensures that the belief qualifies as knowledge. There is simply no need for X's appearing to her to be a bear to convey the information that X is an animal. (v) Intuitions supporting CR+experientiality can be accommodated by distinguishing between requirements for knowledge itself, i.e., generic knowledge, and requirements for an important species of knowledge that can be called *experiential knowledge*. S has *experiential knowledge* that P iff S believes that P on the basis of a conclusive experiential reason. Although Mia's belief that X is an animal qualifies as generic knowledge, it doesn't qualify as experiential knowledge.

6. Skepticism and Closure

We now attempt to show that DAK accounts for the *prima facie* plausibility of closure-invoking skeptical arguments, and nonetheless implies that they are fallacious. Our guiding hypothesis will be that such arguments exploit confusions pertaining to generic knowledge on the one hand, and to various species of knowledge on the other. In some cases, the skeptical arguments presuppose the truth of one of the competitors of DAK we've discussed in this paper.

Ann believes that X is a zebra on the basis of a conclusive reason consisting of X's appearing to her to be a zebra. DAK implies that she knows X is a zebra on the basis of having this experience. She would have this experience were X a *zebra-doppelganger*, i.e., a non-zebra that appears to her to be a zebra because she is a victim of massive deception by an evil genius. However, the situation is such that if X weren't a zebra, it wouldn't be a zebra-doppelganger—it would be an elephant, a giraffe, or something else that wouldn't appear to her to be a zebra.

Knowing that X's being a zebra implies its not being a zebra-doppelganger, she infers that it isn't such a non-zebra. Closure implies that her conclusion belief qualifies as generic knowledge. This is also implied by DAK, epistemic-basing and inferential-reasons. Inferential-reasons implies that she believes X isn't a zebra-doppelganger because she believes it's a zebra, and epistemic-basing implies that she believes it isn't such a non-zebra on the basis of its being a zebra. Since X wouldn't be a zebra if it were a zebra-doppelganger, its being a zebra is a conclusive reason for her belief that it isn't such a non-zebra. DAK implies that her conclusion belief qualifies as generic knowledge.

Suppose that Ann's premise beliefs qualify as defensible knowledge. Neither closure nor the above-mentioned theses imply that her conclusion belief qualifies as defensible knowledge. Indeed, it's plausible that the latter belief is indefensible. She believes X is a zebra on the basis of its appearing to her to be one. She goes on to believe that X isn't a zebra-doppelganger on the basis of it's being a zebra. It's plausible that refraining from doing this might be at least as reasonable from her own perspective as doing it, were she concerned at the time only with acquiring the truth regarding whether or not X is a zebra-doppelganger by doing it. (If she defensibly knew that she defensibly knew that X is a zebra, she might acquire defensible knowledge that X isn't a zebra-doppelganger. In such a case, however, her conclusion belief would be based on her defensibly knowing that X is a zebra rather than on X's being a zebra.) Thus, it's plausible that she lacks defensible knowledge that X isn't a zebra-doppelganger. In many everyday situations beliefs are considered to qualify as knowledge only when they are defensible, and the indefensibility of a given belief is often construed as a good reason to think and to say that it doesn't qualify as knowledge. Consequently, the obvious indefensibility of her conclusion belief may give rise to the mistaken impression that it doesn't qualify as knowledge.

The following closure-invoking skeptical arguments purport to show that Ann's conclusion belief that X isn't a zebra-doppelganger fails to qualify as knowledge, and that, as a consequence, her premise belief that X is a zebra also fails to qualify.

Skeptical Argument 1: Assume for the sake of the argument that Ann's belief that X is a zebra qualifies as knowledge. Closure implies that she knows X isn't a zebra-doppelganger. If X were a zebra-doppelganger, however, she might possess the same reason for believing that it isn't a zebra-doppelganger that she actually possesses, viz. its appearing to her to be a zebra. Hence, she doesn't know that it isn't a zebra-doppelganger. Since the assumption implies a falsehood, it is false.

This argument is fallacious because the assumption doesn't imply a falsehood. Although Ann doesn't know that X isn't a zebra-doppelganger on the basis of its appearing to her to be a zebra, she does know this on the basis of X's being a zebra.

A skeptic who is influenced by the foundationalist portrayal of a subject's fund of knowledge might respond by arguing along the following lines:

Since X's appearing to Ann to be a zebra isn't a knowledge-sufficing reason for her belief that X isn't a zebra-doppelganger, her having this experience doesn't sufficiently support the belief, and therefore doesn't sufficiently support her belief that X is a zebra. Hence, neither of these beliefs qualifies as knowledge. Since DAK implies that these beliefs do qualify as knowledge, DAK should be replaced with CR+unassailability, which is equivalent to: S knows that P iff S unassailably believes that P on the basis of a conclusive reason, i.e., iff S believes that P on the basis of a reason, R, that is such that, for every Q that is implied by P, R wouldn't be the case if it weren't the case that Q. CR+unassailability implies that Ann doesn't know that X isn't a zebra-doppelganger, and implies that she doesn't know that X is a zebra.

The informationalist portrayal of Ann's fund of knowledge can serve to highlight the arbitrary and misleading character of this foundationalist portrayal of it. In virtue of being a conclusive reason for Ann's believing that X is a zebra, X's appearing to her to be a zebra conveys the information that X is a zebra, and enables her believing that X is a zebra to result in her believing that X isn't a zebra-doppelganger on the basis of this information. Since this basis constitutes a conclusive reason for the belief, it conveys the information that X isn't a zebra-doppelganger and ensures that the belief qualifies as knowledge. X's appearing to her to be a zebra, which is a mediate reason for her believing that X isn't a zebra-doppelganger, isn't a conclusive reason for her believing this, and therefore doesn't convey the information that X isn't a zebra-doppelganger. While the foundationalist portrayal of her fund of knowledge suggests that this mediate reason doesn't 'sufficiently support' her belief that X isn't a zebra-doppelganger, the informationalist portrayal indicates that there is simply no need for this mediate reason to convey the information that X isn't a zebra-doppelganger. X's being a zebra is a conclusive reason for this belief, and therefore conveys the information that X isn't a zebra-doppelganger and ensures that the belief qualifies as knowledge. Accordingly, CR+unassailability should be rejected, as it specifies implausibly stringent requirements for knowledge.

Skeptical Argument 2: Assume for the sake of the argument that Ann's belief that X is a zebra qualifies as knowledge. Closure implies that she knows that X isn't a

zebra-doppelganger. Since she knows both that X is a zebra and not a zebra-doppelganger, she is in a position to know that it's zebra rather than a zebra-doppelganger. If it weren't the case that X is a zebra rather than a zebra-doppelganger, however, it might be the case that X is a zebra-doppelganger. And if X were a zebra-doppelganger, it would appear to her to be a zebra. Hence, she is in no position to know that X is a zebra rather than a zebra-doppelganger. Since the assumption implies a falsehood, it is false.

This argument is fallacious because the assumption doesn't imply a falsehood. The argument involves the false presupposition that if Ann knows both that X is a zebra and that X isn't a zebra-doppelganger, she is in a position to know that X is a zebra rather than a zebra-doppelganger. According to contrastive-knowing, she knows that X is a zebra rather than a zebra-doppelganger only if she believes this on the basis of a differentiator that enables her to distinguish between the competing possibilities, X's being a zebra and X's being a zebra-doppelganger. X's appearing to her to be a zebra cannot play the role of a differentiator, for she might have this experience if X were a zebra-doppelganger. And X's being a zebra cannot play this role, for X's being a zebra entails X's not being a zebra-doppelganger. If X's being a zebra could play this role, knowing that X is a zebra *rather than* a zebra-doppelganger couldn't constitute an epistemic achievement that was superior to knowing that X is a zebra and knowing that X isn't a zebra-doppelganger.³⁹ Thus, the assumption doesn't imply the false proposition that she is in a position to know that X is a zebra rather than a zebra-doppelganger.

Skeptical Argument 3: Assume for the sake of the argument that Ann's belief that X is a zebra qualifies as knowledge. Closure implies that she knows that X isn't a zebra-doppelganger. Since her conclusion belief is based on a conclusive reason, viz. X's being a zebra, it would qualify as knowledge were it defensible. But it fails to qualify as knowledge because it is indefensible. Since the assumption implies a falsehood, it is false.

This argument involves a commitment to the truth of CR+defensibility: S knows that P iff S defensibly believes that P on the basis of a conclusive reason. As the following considerations show, acceptance of CR+defensibility is incompatible with acceptance of closure. If CR+defensibility and closure are both true, then the

³⁹ This is not to deny that in some contexts a speaker who utters the sentence "Ann knows that X is a zebra and not a mule" might affirm not only that Ann knows both that X is a zebra and that X isn't a mule, but also that she knows that X is a zebra rather than a mule.

fact that Ann's conclusion belief that X isn't a zebra-doppelganger is indefensible implies that her premise belief that X is a zebra, which is based on a conclusive reason, must be indefensible. But this belief seems perfectly capable of being defensible – it seems quite possible that her believing that X is a zebra on the basis of X's appearing to her to be a zebra would be more reasonable from her own perspective than not doing this, were she concerned at the time only with acquiring the truth regarding whether or not X is a zebra by doing it. A defender of Argument 3 might respond by contending that, even if her belief is initially defensible, it becomes indefensible as soon as she comes to know that X's being a zebra implies its not being a zebra-doppelganger, for she then possesses a good reason to doubt that X is a zebra, viz. the very fact that its being a zebra implies its not being a zebra-doppelganger. This contention, however, is implausible, as it implies that no amount of experiential evidence would render the belief sustainably defensible. Assume for the sake of the argument that: (i) initially, she defensibly believes that X is a zebra on the basis of a conclusive reason consisting of X's appearing to her to be a zebra, X's sounding like a zebra to her, X's smelling like a zebra to her, X's seeming to her to walk and run like a zebra, etc.; and (ii) she then comes to know that X's being a zebra implies X's not being a non-zebra that appears to her to be a zebra, sounds like a zebra to her, smells like a zebra to her, seems to her to walk and run like a zebra, etc. In virtue of being committed to acceptance of CR+defensibility and to closure, the skeptic is committed to acceptance of the implausible proposition that she no longer *defensibly* believes that X is a zebra. It seems clear, then, that acceptance of CR+defensibility is incompatible with acceptance of closure.

Skeptical Argument 4: Assume for the sake of the argument that Ann's belief that X is a zebra qualifies as knowledge. Closure implies that she knows that X isn't a zebra-doppelganger. Her conclusion belief, however, is indefensible. Hence, in virtue of the truth of the following two theses, the assumption is false: (i) Closure, and (ii) Beliefs qualify as knowledge only if they are defensible and non-accidentally true. These theses are among the 'analytically true' principles of a folk theory about non-accidentally true belief—call it *folk epistemics*—that involves the conceptualization of such belief as *knowledge*. The principles of folk epistemics, taken collectively, have the unfortunate consequence that the concept of knowledge is virtually inapplicable to real world situations—very few, if any, non-accidentally true beliefs can qualify as genuine knowledge. While ordinary skeptics have correctly contended that little or no knowledge exists, they have typically neglected to go on to accept the view—call it *metaskepticism*—that the concept of knowledge is part of a fundamentally flawed folk theory that should be replaced with one that more accurately represents the nature of non-accidentally true belief. Such a replacement theory

could serve as a highly useful guide for philosophers and cognitive scientists investigating the important role that non-accidentally true belief plays in human life.

This argument presupposes that the concept of knowledge is analogous to the ‘theoretic’ concepts that constitute integral components of many theories. Furthermore, the argument involves the contention that closure-invoking skeptical arguments are sound. If such arguments turn out to be sound, metaskepticism may turn out to be an attractive version of skepticism, provided that the concept of knowledge can be convincingly construed as a ‘theoretic’ concept. As we have shown above, however, there are good reasons to think that closure-invoking skeptical arguments are unsound. Epistemic-basing, DAK, and inferential-reasons are plausible theses that discredit closure-based skepticism. Hence, there are good reasons to think that metaskepticism is unacceptable.

7. Conclusion

We have argued that DAK not only accords with closure, but also implies it. In addition, we have argued that DAK accounts for the *prima facie* plausibility of closure-invoking skeptical arguments, and nonetheless implies that they are fallacious. If our arguments turn out to be sound, the acceptability of DAK will be significantly enhanced by the fact that, despite implying closure, it undermines skepticism. While we think that DAK is superior to each of the competing analyses we have discussed, we end this paper with an admission that three of these competitors are, in our opinion, still in the running: CR+experientiality, CR+contrastivity, and CR+defensibility. The first excels at accommodating foundationalist intuitions, the second at accommodating contrastivist intuitions, and the third at accommodating internalist intuitions. Like DAK, these analyses undermine skepticism; but they do so by falsifying closure. It seems clear, therefore, that if one deems acceptance of closure and rejection of skepticism to be of overriding importance, one should seriously consider accepting Dretske’s analysis of knowing.⁴⁰

⁴⁰ We are deeply grateful to Fred Dretske for providing us with very helpful comments on earlier versions of this paper. In addition, we have benefited greatly from discussions with Alvin Goldman, Robert Gordon, John Hyman, Peter Klein, William Larkin, Thomas Paxson, Gualtiero Piccinini, Ernest Sosa, and James Stone.

SCIENTIFIC RATIONALITY AS NORMATIVE SYSTEM

Vihren BOUZOV

ABSTRACT: Decision-theoretic approach and a nonlinguistic theory of norms are applied in the paper in an attempt to explain the nature of scientific rationality. It is considered as a normative system accepted by scientific community. When we say that a certain action is rational, we express a speaker's acceptance of some norms concerning a definite action. Scientists can choose according to epistemic utility or other rules and values, which themselves have a variable nature. Rationality can be identified with a decision to accept a norm. This type of decision cannot be reduced only to its linguistic formulation; it is an act of involvement of the normative regulation of human behavior. Norms are treated as decisions of a normative authority: a specific scientific community is the normative authority in science. These norms form a system and they are absolutely objective in the context of individual scientists. There exists an invariant core in all the norms of rationality, accounting for their not being liable to change, as compared with the flexibility of legal norms. The acceptance of and abidance by these norms is of social importance—it affects the aims of the community. A norm only defines the common framework and principles of scientific problem-solving; its application is a matter of professional skills and creative approach to a particular problem. It is of no importance at all, if an agent's cognitive abilities do not live up to the requirements of a norm. Such discrepancy can be compensated for by the fact that a scientist carries out work in a conceptual and normative framework established by a respective scientific community.

KEYWORDS: norm, decision, normative system, scientific community, scientific rationality

1. On the Pluralism of Rationality

In my view, it would not seem to be warrantable to take in science as a pattern of rationality, and to consider scientific activity as being more rational than other types of human activity, without a clear-cut understanding of the concept of rationality, scientific rationality in particular. On the other hand, its perceiving of as an abstract construction with dogmatic and restrictive characteristics—or as a wholly evaluative concept void of content—would have a serious grounding.

There exists no agreement in the overwhelming majority of contemporary philosophers about the nature of scientific rationality and its traits. Obviously, it is of major importance to find out positive solutions to problems of the nature of rationality in the context of the intellectual crisis holding sway, when criticism of science and irrationalism are in aggressive offensive and there is talk about ‘collapse’ of scientism, scientific attitude, foundationalist programs in philosophy, and scientific or rationalistic perception of the world.¹ Below I try to explain the nature of this issue by means of referring to the concepts of *norm* and *decision*. Scientific rationality is perceived by me – as a normative system accepted by a specific scientific community. An informal decision-theoretic approach is the methodological instrument of the analysis made by me.

The concept of rationality relates to the instruments of carrying out human activity and defining suitability in terms of aims. The ideas of ‘rationality’ and ‘rational-irrational’ have a philosophical history of their own. Classical philosophical tradition draws a line of demarcation between rationality of thinking and rationality of action, between theory and practice. It is based on a response to the so-called problem of the genesis of knowledge: the main part in it is played by Reason via innate universal knowledge (the so-called ‘innate’ ideas). It is a response of rationalism—the foundation of the so-called ‘modernistic project’ of the Enlightenment, which defines the universal laws of Reason, guiding nature, society, humans and knowledge.² This type of rationality is selfsame for all people and is not dependent on time and social conditions. It characterizes the development of thinking, not that of reality. Rationality of thinking is an emanation of transcendental Reason. Typically, it is identified with the laws of logic and other ‘innate’ truths. The rationality of an action is determined by aspects of: situation of choice, limited ability and knowledge of a given individual, and his free will. These aspects are rational, falling in with aims, and conducive to their realization.

Another conception of the nature of rationality, featured below, is the methodological one: we can think of rationality of science as a definite set of characteristic features of a scientific method. The positivist and postpositivist philosophy of science identify rationality via a set of methodological rules. This conception of rationality presupposes evolvment and availability of a universal

¹ Raimo Tuomela, “Science, Protoscience and Pseudoscience,” in *Rational Changes in Science*, eds. Joseph C. Pitt and Marcello Perra (Dordrecht: Reidel, 1987), 93.

² Stephen Toulmin, *Cosmopolis. The Hidden Agenda of Modernity* (New York: The Free Press, 1990).

method and systematization of sciences. Scientific theories have to abide by certain rules and standards, themselves the gist of logical stringency. Rationality is guaranteed by means of abidance by such rules and standards, themselves an expression of procedures of acceptance, justification and criticism of knowledge. Their uniqueness and logical power determine the priority of science as regards other forms of knowledge. They are means of gaining objective, genuine knowledge; or of adequate explanation of phenomena. Their fathoming leads to the construction of rational models with claims on ability to reveal the nature of scientific knowledge and scientific change (Here I mean the models brought out by Karl Popper, Thomas Kuhn, Imre Lakatos, Larry Laudan, Paul Feyerabend and others). In this methodological context we can understand the definitive words of Kazimierz Ajdukiewicz, the prominent polish philosopher, who treats rational knowledge as “intersubjectively communicable and verifiable” by means of the use of objective methods.³

Are scientists rational in terms of the “methodological conception of rationality?” Lars Bergström is right in saying that it “confuses means and ends, or process and product, in a certain way.”⁴ Methodological rules could be perceived of as forwarding some of the aims of science, not as determining any particular behavior of individual scientists.

So far, the contemporary philosophy of science has not been successful in proving, convincingly, that rationality of scientific knowledge might be perceived of as one keeping up to rigid methodological rules. Paul Feyerabend thinks that such type of rationality is a holdback in the feasible advance in science; it imposes limitations on human freedom. Scientific progress makes headway through breaking up the constraint of methodological rules.⁵ The hope that such general and all-embracing directives exist has been dwindling away all along, primarily due to the impact of the established pluralism of forms of rationality. Feyerabend convincingly points to the real variety of ‘rational’ standards. The latter determine different cognitive strategies and practices. One might rightfully infer that the interpretation of a certain cognitive procedure or action as rational ones could not be pared down to a finite set of qualities and characteristic features. The concept

³ Kazimierz Ajdukiewicz, *Zagadnienia i kierunki filozofii. Teoria poznania. Metafizyka* (*Problems and Theories of Philosophy. Theory of Knowledge. Metaphysics*) (Diamonion, 2003), 50 (in Polish).

⁴ Lars Bergström, “Some Remarks Concerning Rationality in Science,” in *Rationality in Science*, ed. Risto Hilpinen (Dordrecht: Reidel, 1980), 1-3.

⁵ Paul Feyerabend, *Against Method. Outline of One Anarchist Theory of Knowledge* (London: Verso, 1975).

of rationality is of a relative and changeable nature. There exists no idea (or activity, tradition) that might be assessed as „the one-and-only rational”, for good. Richard Rorty works out to an extreme relativism this entirely grounded conclusion of Feyerabend to an extreme relativism.⁶ This assertion is entirely unacceptable, because there are also evaluative and normative invariants going to the making of rationality.

At present, following pragmatist criticism—and on the basis of the use of the decision-theoretic approach—philosophy stipulates an elimination of the difference between thought and action, and between theory and practice. Thought is considered to be a type of practical activity, a singling out of alternative decisions. A subject's development is a process, the nature of which is determined by internal and external factors. We can say that the distinction made between methodological and practical rationality, between inferential and behaviorist conceptions of knowledge and reasoning, arises out of the unjustifiable 'thought/action opposition.'

The decision theory is, as I see it, the most successful winner in the evolvment of a model of practical rationality. In terms of practice, rationality is a choice padded with good grounds.⁷ The theory of decision seeks to offer a plausible model of rational action and to formulate general principles of rationality, guiding decision-makers under conditions involving risk and unreliability of information. 'The agent' has to make a choice in the presence of several alternatives: their results depend on the actual occurrence of a situation—reciprocally excluding each other in a set of situations. The agent will be striving to act in a way that might bring about a maximum meeting of his needs or preferences. A choice is rational if it maximizes an expected utility (usefulness) of a given action. This is the main principle of rationality in the theory of decision.

There exists a shared agreement that the decision theory can be applied to the problem of scientific rationality, 'hard problems' in particular, such as confirmation and justification.⁸ A relatively successful methodological trend in the contemporary philosophy of science is developed on the basis of decision logic. One could speak about epistemic utility and about choice of scientific hypotheses as an activity modeled by some rules of rational decision-making. However, this

⁶ Richard Rorty, *Philosophy and the Mirror of Nature* (Oxford: Blackwell, 1980), 331.

⁷ Frederic Shick, *Making Choices. A Recasting of Decision Theory* (Cambridge: Cambridge University Press, 1997), 34.

⁸ Ronald N. Giere, *Understanding Scientific Reasoning*, 4th edition (Orlando: Harcourt Brace College Publishers, 1997), Colin Howson and Peter Urbach, *Scientific reasoning: The Bayesian Approach* (La Salle: Open Court, 1989).

new methodological paradigm calls for existence of a more convincing conceptual justification based on the concepts of norm and normative systems as regards human action and its normative regulation.⁹

2. Types of Rationality of Action

It is true that all typologies of rationality of action are based on the making of a distinction between rationality of ends and rationality of means for the fulfillment of aims. They might be defined as *axiological* and *instrumental* types of rationality. It is a characteristic feature of European thought that it interprets reality by means of the use of models in the context of the ‘means-ends’ relationship, yet. Models are abstract conceptual structures representing the main characteristics of reality.

Instrumental rationality can be termed as technological or economical one, too. We can consider it as a choice of means in the realization of a definite end through minimal effort. Their ‘ratio’ is a yardstick of action effectiveness. Instrumental rationality encompasses the real essence of the capitalist organization of society and of its bureaucratic administration and economy.

Instrumental rationality has different forms of manifestation. As regards organizations it functions as *system rationality*, featuring the need for of effective implementation of definite organizational objectives. It can also be defined as *action rationality*—in the context of the practical situation of making a choice of alternatives.

In his paper *Rationality as a Value* Klemens Szaniawski, another prominent Polish philosopher, emphasizes that rationality is a “fully rational value, which has positive or negative meaning as regards respective aims.”¹⁰ Axiological rationality is determined by a choice of appropriate aims. The task of formalizing the axiological content of a decision is very difficult. Choice of aims is determined by: value orientation, subjective preferences and empirical experience. The definition of an aim is an objective realization of thought. If one wants to fathom the process of discovery, formulation and realization of aims—he has to get to know the essence of a thought in its relations to reality. Here one has to deal with scientific rationality—with scientists’ search for realization of some scientific aims.

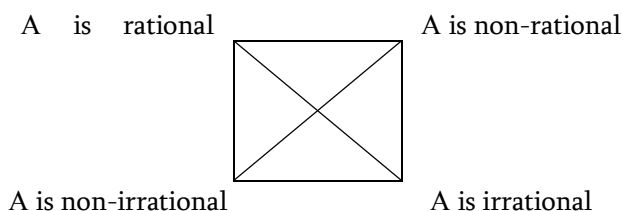
⁹ Vihren Bouzov, “Scientific Rationality, Decision and Choice,” in *Bulgarian Studies in the Philosophy of Science, Boston Studies in the Philosophy of Science*, Volume 236, ed. Dimitri Ginev (Dordrecht: Kluwer, 2003), 17-29.

¹⁰ Klemens Szaniawski, “Rationality as a Value,” in his *On Science, Inference, Information and Decision-Making: Selected Essays In The Philosophy Of Science*, eds. Adam Chimelewski and Jan Wolenski (Dordrecht: Kluwer, 1998), 232-240.

The opposition between *formal* and *cultural* rationality is another aspect of our principal classification of rationality. Formal rationality presupposes availability of certain objective criteria and measures of choice-making, all of them with a quantitative expression (Example: an individual's choice of some marketed goods). Cultural rationality is determined by selection of aims: it has an evaluative basis rooting in cultural, social and individual experience.

Referring to the use of old philosophical approaches, we can distinguish between subjective and objective rationality. Rationality, as an evaluation, expresses acceptance by a given evaluator of specific norms determining an agent's behavior.¹¹ The *evaluator* can be an individual, a social group or a society (here I mean also self-evaluation). The objective content of rationality spells out the relation between an action and a state of the world. The rationality of science itself has an objective aspect, too.

Rationality can be considered as modality, as well.¹² We can interpret the context of "A is rational" in this way. In such context, with variable A means beliefs interpreted as epistemic relations to propositions. It can be applied to descriptions of actions. Rationality cannot be reduced to the definition of truth. Referring to an analogy with the classical *logical square*, we can expand the area of possible rational evaluation of human action as follows (A is a proposition or an action-description):



All relations—in the well-known logical square are intact—contraries are mutually-excluded, subcontraries are mutually-added, diagonal ones are in contradiction, subalterns propositions are in a relation of logical consequence—from general to particular. We can deduce "A is non-irrational" from "A is rational" and "A is irrational" from "A is non-rational." Rationality is opposed to irrationality, non-rationality is opposed to non-irrationality.

¹¹ Bouzov, "Scientific Rationality."

¹² Jan Wolenski, "Racjonalność jako modalność" ("Rationality as Modality"), in *W stronę logiki* (*From the Point of View of Logic*) (Krakow: Aureus, 1996), 125-137 (in Polish).

This typology directs us to getting over contradiction between rationality and irrationality. An action can be non-rational, not irrational (for example: buying a present for a girlfriend). Scientific activities can be rational or non-rational.

3. Norms, Values, and Scientific Rationality

When we say that a certain action is rational we do not express an assertion, be it true or false. This type of assertion has a definite comparative and evaluative element: we express a speaker's acceptance of some norms permitting or prohibiting the performance of an action. According to Isaac Levi, the requirements of coherence and consistency—the so-called “weak principles of rationality”—are “normative standards of rational health.” They “could be deployed by deliberating agents to evaluate their options, probability judgments and value judgments;” they should be applicable to self-criticism as well.¹³ Hence, it is an “action-guiding dimension” of rationality; in this sense the decision theory, as a normative theory, “provides normative criteria for assessing how decision problems are resolved.”¹⁴

But “the external perspective” is of greater importance in rational evaluation. The decision theory can be a methodological tool in predicting or explaining human behavior. Norms themselves are *decisions* of a normative authority. The value judgment of a respective scientific community is external to an individual scientist. The external perspective of the application of rational normative standards is objective and is determined by social factors. Statements of rationality can be objective even if ‘the rational’ is only valid for humans in specific contexts.¹⁵ Therefore, one can say that scientific rationality is a non-stringent regulatory system.

Norms are prescriptions for action, based on values and systems of preferences, yet, they are of an objective nature, too. They are introduced by performative utterances of the type of: “I state that A is obligatory (prohibited),” thus expressing a decision of a certain normative authority. The formulation of

¹³ Isaac Levi, *The Covenant of Reason. Rationality and the Commitments of Thought* (Cambridge: Cambridge University Press, 1997), 24-6.

¹⁴ José Luis Bermúdez, *Decision Theory and Rationality* (Oxford: Oxford University Press, 2009), 14.

¹⁵ Hillary Putnam, “Pragmatyzm i wiedza pozanaukowa” (“Pragmatism and Extrascientific Knowledge”), in *Pragmatyzm i filozofia Hilarego Putnama (Pragmatism and the Philosophy of Hilary Putnam)*, ed. Urszula M. Zeglen (Torun: UMK, 2001), 24-5 (in Polish).

norms bears on 'the will,' but it is not devoid of rational grounds. A decision cannot be reduced to its linguistic formulation only; saying that it is an act of evolvment of normative regulation, a process of imposing an authoritative will, the result of which is a division of all possible actions into three, mutually-disjoint sets: obligatory, forbidden and indifferent. The decision to enact a norm and its acceptance by an addressee are actions. Norms are ordered pairs of the type of $\langle OA_i, K_i \rangle$, where $OA_i \in X$, X is a set of initial obligations $\{OA_1, \dots, OA_m\}$, the variable A_i expresses actions, and for every $W \in K_i$ (the set of possible worlds), $V(A_i, W) = 1$. In the context of logic, 'normation' spells out choice of a normative function. It is a choice of a set of postulated possible worlds imposed on the real world (the *normed Universe*) as its deontic alternatives. Obligations and prohibitions, established in the real world, are realized therein.

Normativity (and, in general, directivity) is viewed as a pragmatic, not as a semantic characteristic of utterance; it cannot be pared down to the concept of 'truth in model.' Normative discourse is made up of deontic propositions and performatives—themselves of a propositional character. The conception of norms, developed above, includes some basic ideas of the so-called non-linguistic theory of norm, suggested by Jan Wolenski.¹⁶ I think that this theory does not give answers to important epistemological questions—it could only be instrumental in asking questions about verification and justification of norms.¹⁷

The logic of norms can be bolstered up in a broader context of decision logic, which has a prescriptive force. As stated above, it is a normative theory in nature. The choice of a norm is a rational choice based on definite rules; they are not entirely formal ones, because they have a definite social content. On this basis, decision logic can be specified as an *intensional* logic of rational choice. Decision-makers can conceptualize the situation in which their choice is made. It can be expressed by means of choosing a suitable norm, acceptable to a given community.

A norm only defines the common framework and basic principles of scientific problem-solving. Normative decision is a choice called upon to substantiate aims of a normative authority. Scientific community is the authority in science. It enacts and guarantees the binding force of the norms of scientific rationality, on the basis of knowledge, empirical verification, tradition, general understanding of science and its aims, taking into account existing social factors

¹⁶ Jan Wolenski, *Z zagadnień analitycznej filozofii prawa (Some Problems of the Analytical Philosophy of Law)* (Warszawa-Krakow: PWN, 1980), ch. III. (in Polish), Jan Wolenski, "Deontic Sentences, Possible Worlds and Norms," *Reports on Philosophy* 6 (1982): 65-73.

¹⁷ Vihren Bouzov, "Norms as Decisions," *ARHE, Casopis za filozofii* 2 (2004): 113-7.

and prescriptions. These norms are absolutely objective in the context of work done by individual scientists or members of a scientific thought collective. Their violation dooms a scientist's research to failure. But this fact does not question the statute of norms. We can agree that scientific rationality is not "a code of directives, applied mechanically, but is a set of general directions with ethical value."¹⁸ The acceptance of and abidance by these norms is of social importance – it affects community aims. The specific content of systems of rationality norms and their historical development is determined by different cognitive and social factors of variable nature. The norms of Aristotle science are different from the norms of modern science; they themselves are subject to change nowadays.

Following suitable analogy with laws of nature, we can say that norms are relatively independent of normative decisions. An agent makes a decision to accept or reject rational norms, because such decisions are in chime with his own interpretation of science's aims and problems, and with the interpretation accepted by the scientific community he belongs to. The choice of a norm is determined by the interest taken in maximizing an expected 'epistemic utility' (Carl Hempel). It is important to emphasize that the interpretation of norms developed by me is not a form of theories of 'norm conformity'.¹⁹ A norm only defines the common framework and the principles of scientific problem-solving; its application is a matter of professional skills and creative approach to a particular problem. It presents possible "good grounds" determining rational choice.²⁰ One can say that it is not justifiable to compare the role of norms of scientific rationality and their collective acceptance with religious fate²¹—they include requirements of criticism and free choice; and they are of importance in the creative process.

It is of no importance at all, if an agent's cognitive abilities do not live up to the requirements of a norm. Such discrepancy can be compensated for by the fact that a scientist carries out work in a conceptual and normative framework established by a respective scientific community. There exists an invariant core in all the norms of rationality, accounting for their not being liable to change, as compared with the flexibility of legal norms. The requirements of empirical and

¹⁸ Szaniawski, "Racionalnosc," 536.

¹⁹ Steven Hetcher, *Norms in a Wired World* (Cambridge: Cambridge University Press, 2004), 4.

²⁰ Shick, *Making Choices*, 34.

²¹ Wang Shan Bo, "The Link Between Scientific Rationality and Religious Rationality," *Journal of Dialectics of Nature* 28, 4 (2006).

theoretical justification of knowledge,²² critical attitude, explanatory and predictive force, can play such a role.

Such a role and the development of various systems of norms of scientific rationality (in the synchronic and diachronic aspects) might be a subject of another philosophical analysis.

²² Peter Maher, *Betting on Theories* (Cambridge: Cambridge University Press, 1993), 25-30.

TWO FALLACIES

James CARGILE

ABSTRACT: In charging *argumentum ad hominem*, we accuse someone of attacking the source of a claim. In charging *argumentum ad verecundiam*, we attack the source of a claim. This is reason for attending to “attacking the source.” It is important to distinguish probabilistic reasons for doubting a claim and evidentiary reasons. Evidence that the source of a claim is likely to be wrong is not evidence against the claim. The tendency to overlook this is the essential feature of the *ad hominem* fallacy. This is relevant in assessing the view that someone who regards his thinking as made possible by Godless arrangements of matter largely determined by chance is, in taking this attitude, advancing a hypothesis which undermines his theorizing about the world or himself.

KEYWORDS: probabilistic reasons, evidentiary reasons,
motivational reasons

There is considerable latitude in descriptions of the fallacy of *argumentum ad hominem*. One version makes it simply a matter of attacking the source of a disputed claim rather than addressing the claim itself. There is nothing generically wrong with this, and anyway, the attack may not be offered as a basis for a conclusion. Digressing to denounce an opponent or source of an opposing view is something which may be done in an argument, but is not a pattern of argument, any more than interrupting the argument to rest and read the newspaper is. A more specific kind of *ad hominem* would be to offer a criticism of your opponent as a basis for rejecting his view. You might advance the claim that your opponent is a dishonest politician in support of rejecting his claim that there are no dishonest politicians in his party. This case would not be one of a fallacy of relevance. In a case where the criticism is not relevant to your conclusion, that would be bad, and it is perhaps worth warning against the temptation to argue in this way.

A still more specific kind of *ad hominem* involves concluding that P is false or unjustified on the grounds that it was put forward by a defective source. To find that a case of this is fallacious is to conclude that the arguer has not justified his conclusion that P is false or unjustified. The fallacy of *argumentum ad verecundiam* involves concluding that P is true or justified on the grounds that it is endorsed by a reliable authority, where in fact, the authority is defective, or the

arguer does not have adequate basis for believing the authority to be reliable. (So this is a premise fallacy rather than an inferential one.) To find someone guilty of this fallacy is to conclude that he has not justified his conclusion that P because the source for P to which he appeals is defective or not justifiably believed reliable.

Accusing someone of a fallacious appeal to authority thus bears considerable resemblance to arguing *ad hominem*, in that it attacks the source of a claim as a basis for finding the claim unjustified. It is worth clarifying the relation between these two fallacies. It would be unsatisfactory if finding someone to have committed the fallacy of appeal to a bad authority was itself to commit the *ad hominem* fallacy. Consider these two patterns:

Pattern I: A claims that P. The question whether P is a kind with respect to which A is untrustworthy (unreliable, insane, etc.). Therefore, A's claim is unjustified.

Pattern II: A claims that P solely on the grounds that B testifies that P. A is not justified in regarding B as a reliable authority. Therefore, A's claim is unjustified.

These patterns have the same conclusion, and in the principal case, both support that conclusion by criticizing the reliability of a source of the claim. (II also includes a secondary case in which the objection is that the arguer is not justified in regarding B as a reliable authority. We will set that case aside in the following discussion.) In II, the source may be a different person B. But in the case where A=B, that distinction between I and II is eliminated. Still, even in that case, there must be a good distinction between I and II, since I is a characteristic pattern for the *ad hominem* fallacy, while II is delivering a verdict of fallacious appeal to authority. I is of course not generically fallacious and it is worth making clearer the conditions under which instances of it are fallacious, but those instances should not include pattern II. It should not be generically fallacious to find someone guilty of a fallacy. It is thus worth explaining why I is not a generically reliable style of reasoning and is such that to treat it as a reliable pattern would be to commit the *ad hominem* fallacy, while II is acceptable.

To begin with II, it is a good pattern only when understood with certain qualifications. If A *offers* only the consideration that P was endorsed by B, he may nonetheless have other adequate reason for claiming that P. Boyd may defend his claim that Bob did it by appeal to the fact that Bill testifies that he saw Bob do it. Bill may be a notorious liar, so that Boyd has failed to justify his claim. But Boyd may himself have seen Bob do it. He may have selected the appeal to Bill's authority out of a very unwise assessment of Bill's authority, thinking that his

hearers will be more impressed by hearing it comes from Bill. Boyd has failed to justify his claim, but he *has* adequate justification for making it.

This qualification cannot be adequately included by changing from “based solely on the grounds that...” to “having as grounds, only the consideration that...” This latter is too unclear. Boyd saw Bob do it. Boyd also knows that his eyes are in good working order, and that he is not prone to wild distortions of memory, etc. He does not mention these considerations in offering justification for his claim that Bob did it. He should have mentioned that he saw it. But to claim that his eyes are reliable is best left unsaid. Still it is a consideration. Just where such considerations leave off would be hard to say, and worthless to attempt to describe. The change would make pattern II much less practically applicable than it should be.

In a given context, the presentation of certain premises $P_1, P_2, \dots P_n$ may be adequate to justify concluding that C . One might think that the contextual features which make $P_1 \dots n$ adequate could be summarized in an additional premise P_{n+1} , to yield a valid and context-independent justification. This is mistaken. A premise can always be found, the addition of which makes the argument valid. All we need is “If $P_1 \dots n$, then C .” Whenever anyone argues “ $P_1 \dots n$; therefore, C ,” if he is sincere, he believes that *if* $P_1 \dots n$, *then* C . If we just add that which he presumably accepts anyway, we always get formal validity. But there is no general guarantee that that will not beg the question.

If it is granted that the features of a situation which make a conclusion reasonable cannot always be stated to yield a context independent argument to justify the conclusion, then we should see that what a reasoner has as grounds or basis for the reasonableness of his conclusion cannot in general be stated completely. Thus we will not be in a position to say that all someone has to ground his claim is such-and-such authority. By contrast, it is easy to be clear about the fact that all someone has offered to ground his claim is such an appeal.

We could change the conclusion from “A's claim is unjustified” to “A has not succeeded in justifying his claim.” This latter is a possible interpretation of the original words, and it makes pattern II quite solid. If A has based his case for P solely on a bad authority, and has offered nothing else (such as good reason to justify his (nonetheless mistaken) assessment of the authority), then he has not succeeded in justifying his claim, whether or not he is in fact justified in making the claim.

This alteration (or clarification) of the conclusion would not help pattern I. It is just as bad to conclude from the premises of I that A has failed to justify his

claim as to conclude that his claim is unjustified. It is part of our present project to explain this.

We have two versions of the conclusion: (i) "A has not presented adequate justification for his claim" and (ii) "A does not have adequate justification for his claim." I propose to understand (i) in such a way that it follows from (ii), but not conversely. This interpretation is not the only possibility. We can understand saying that someone presented adequate justification for a claim without his being justified in that claim. Someone might present Q,R and S, and these be adequate justification for P, without his getting credited with *having* adequate justification for P, if he himself did not believe Q,R or S and was presenting them with the intent to deceive.

We could understand saying that someone had adequate justification for asserting that P but was not justified in asserting that P. He might know certain things which justify concluding that P and yet fail to recognize this and not believe and thus not assert that P. Thus he 'has' justification in one sense, but not in another. Furthermore, just adding belief that P would not be enough. It would have to get properly connected with the justifying considerations. I will take having adequate justification to entail being justified. More important here is the notion of successfully presenting adequate justification. As I use this notion, it entails having adequate justification. This is not at all strictly observed in ordinary usage, but hopefully it will be clear enough in what follows.

The possibility of an alternate understanding on which a presenter of adequate justification may not himself have adequate justification may explain why people tend to become more wary when presented with Pattern Ii than when given Iii. A thoroughly untrustworthy person may be paid to 'present' the justification of a reliable authority. We may not credit the presenter with being justified in making any of the claims involved while still crediting them to the person for whom he makes the presentation. It seems obvious that merely getting an unreliable person to recite an argument should not serve to make the argument inadequate.

If someone is sufficiently unreliable, it may be reasonable to conclude that he is not justified in claiming that P, for some suitably complex P. Thus if a difficult proposition of mathematics is put forward, and the question is asked whether it is true, if the village idiot pipes up from the back "It is true," we may conclude that this is at best a lucky guess and the man does not even understand the proposition. It could not be true that someone is an idiot (as opposed to, among other possibilities, an *idiot savant*) and yet understood a complex proposition of mathematics. We can reduce this to a matter of how we interpret

the relevant terms. There is a level of incompetence such that it is not possible to be at that level and accomplish certain intellectual successes, just as it is impossible for an absolutely failing student to pass a test. Passing proves conclusively that he is not 'absolutely' failing.

The argument scheme I': "A is incompetent on questions like P, therefore, A did not present a good argument for P such that he can be credited with thus justifying the conclusion that P," is a close relative of Pattern I. Pattern I would infer from the fact that A is incompetent and an argument for P is creditable to A, that the argument is not a good one. I' infers from the fact that A is incompetent that if he has offered a good argument for P, then it is not creditable to him. Another variation (I'') would be to infer that since A has presented a good argument for P, which is creditable to him, he is not incompetent on the topic. To show that I is fallacious involves showing that these variations are also wrong. That is not to show that all instances are bad arguments, but just that the pattern does not give any support to an instance, so that to rely on the pattern is fallacious.

Suppose that the putative village idiot submits a paper to a journal, and the journal politely declines, without reading the paper. When challenged by a peer, they explain that the 'author' is an idiot incapable of producing anything but garble, and they do not have time to read all the papers submitted. This could be a reasonable response. Suppose further, that the peer somehow manages to persuade the editors to read the paper, and they find it contains a brilliant argument for a novel conclusion. They may still quite reasonably object to publishing the article under the idiot's name, on the grounds that he could not have written it. Suppose still further, that the idiot comes in to the office and writes a new paper under close observation, which is also found to contain a brilliant and original argument. Then there is no course but to withdraw the claim that the man is an idiot, or totally incompetent. This still does not undermine the general correctness of the inference from "He's an idiot" to "His paper is worthless." It merely shows the premise of that inference to be false in this case.

Suppose that the putative idiot produces a paper arguing that he is mentally completely incompetent. Again, this is a brilliantly argued paper, appealing to a highly impressive study of the physiology of his brain and the intellectual limitations which follow by well established brain science, from having a brain in such a condition. We are ready to conclude that the production of this paper refutes its thesis. Even though the thesis is wrong, the paper is clever enough to be proof that the thesis is wrong. The paper, of course, does not *contain* proof it is wrong. Rather, the fact it is so clever disproves the thesis that its author is completely incompetent.

The purported idiot may have anticipated this line of criticism and attempted to answer it in his paper. He argues that it is not impossible and not of probability zero that he should produce a cogent argument. It is just fantastically unlikely. We assume the man's general behavior manifests thoroughgoing mental incompetence, so that it is hard to credit his present cogency to the same person. That he is himself amazed, on reviewing what he has written, to think that someone as stupid as he is could have produced such stuff, is a kind of perception, accurate or not, which is not in keeping with his usual manifest character. He says that even though he is too incompetent to be a reliable judge of the soundness of his paper, it looks good to him, and he would like to know what is wrong with it. As he is currently present to us, he is a partner in dialogue, with a claim on our attention and response he would not usually be capable of making.

One reply might be based on the principle that if a theory entails that it is fantastically unlikely that *e* will occur, then the occurrence of *e* is strong evidence against the theory. The idiot's theory implies that it is fantastically unlikely that he would produce a clever paper. But he did produce one, so this theory is refuted. This is a poor reply, since the idiot's theory includes the claim that *e* has happened in spite of being extremely unlikely. Unless you can give reason for thinking there is a better explanation than mere chance, it is question-begging just to complain about the low probability.

Consider this theory *T*:(1) Jones tossed an unbiased coin ten times on occasion *O* and got all heads. (2) (1) was due to chance and had a probability of 0.0009765625. Should we say that the probability of *T*(1) given *T* ($\text{Prob}(T(1)/T)$) is 1, on the grounds that *T* entails *T*(1), or should we say that it is 0.0009765625, on the grounds that *T* entails that is the unconditional probability of *T*(1)? However this is answered, we should recognize that it is question-begging to reject *T* on the grounds that, since one part of it has a probability of 0.0009765625, it is very unlikely that it be correct.

An alternative to *T* is *T'*: that Jones was cheating and was not tossing in an unbiased fashion. This could of course be right, but this depends on further details. To say *a priori* that *T'* is a more likely theory than *T* is unwarranted. It is a feature of our human condition that we recognize that there is a nonzero probability that an unbiased coin should be tossed ten thousand times and land heads every time, and yet we must concede that we could not know this happened. If we observe ten thousand heads, we cannot reasonably ascribe this to chance, even if we can think of no other explanation. We cannot reasonably rule out the hypothesis that something is influencing that coin. If someone claims to have psychic power over a coin and offers to prove it by letting us toss the coin four times which he will

make straight heads, and we do get that result, then it would be reasonable to discount this performance as lucky. If we tossed a thousand in this situation, we would have to admit that there is some power at work. The difference between four and a thousand has to do with background information or inclination.

We do, I believe, have background information (or inclination) such that the production of a clever theory concluding that he is an idiot would refute the theorist. Inability to articulate this background information could lead to the illusion that we can properly reject the theory purely because it has a low probability of being true. But it is incorrect to argue "P is a very unlikely reason why e occurred; therefore it is not true that P is the reason why e occurred." The performance of a purported idiot who writes a brilliant argument for the conclusion that he is incompetent is not self-refuting in the sense of the performance entailing the falsity of its conclusion (though it does provide conclusive reason to reject its conclusion). By contrast, asserting you cannot make an assertion is self-refuting in that way. It is impossible for the premise to be true and the conclusion false.

If the conclusion of the argument is that the arguer is incapable of producing a good argument, then the goodness of the argument would refute the argument. Granting that a self-refuting argument could not be a good argument (which is not obvious, but conceded for purpose of argument), then that argument cannot be a good one. Thus the form: A is incompetent in the sense of being utterly incapable of producing a good argument; A argued that P; therefore, A's argument for P is not a good one; is valid. It is a relative of pattern I which is not an inferential fallacy, though the premise is likely to beg the question in a particular case. But the form: A is incompetent, that is, very unlikely to argue well or get anything right. A argued that P; therefore A's argument for P is not a good one, is not valid. To think otherwise would be fallacious.

Someone may of course, be unreliable on one sort of topic while being reliable on another. The interesting cases are those in which the arguer or assertor is unreliable on the topic about which he is arguing or asserting. Here it is important not to slip into a formulation of 'incompetence' such that the mere performance of presenting the argument proves the arguer is not at that level of incompetence. Such a formulation is worth noting, but is not the most interesting case. Once we are clear of those extreme readings of 'incompetent' it is fairly obvious that from the fact that the arguer is ever so unlikely to be right it does not follow logically that he is not right. What is not so clear is whether it is justified to conclude that he is not right.

Suppose that we are given that an urn contains the numbers of 10,001 propositions. 10,000 are proven falsehoods. One is not known to be false and not known to be true (but it is a definite proposition and thus either true or false). We are not given the content of any of these propositions. They may be about topics such as the doings of people in a foreign country, so that we know nothing about any of them just from hearing the content. We make a random drawing and get proposition number *n*. It is surely reasonable for us to infer that prop. *n* is false. It would be reasonable to infer also, that if a certain stranger believes prop. *n*, then he is mistaken. (We might add the assumption that for each of the 10,000 proven falsehoods, there is someone who believes it.)

Proceeding to the question whether it would be reasonable to infer that the stranger's belief in prop.*n* is unjustified, we should become more cautious. We have not heard his argument, after all. Even if mistaken, he might still be justified. But we can bring this under probabilistic considerations by the same method. We assume an urn of 10,000 bad arguments and 1 unanalyzed argument, and we are given that argument *X* has been drawn randomly from that urn. Then it seems to be highly reasonable to conclude that *X* is a bad argument.

One difference here from the case of unargued (contingent) propositions is that we can evaluate some arguments even when we cannot evaluate the constituent propositions. We may have no idea whether it is true that Wong grows melons in Daigan or that all melon growers in Daigan are in debt, but we can readily see that it is correct to infer from these premises that Wong is in debt. To that extent we can determine the merit of the argument with less dependence on background information.

This still does not let us determine whether the argument is a good one. As an attempt to establish that Wong is in debt, the arguer may have appealed to two egregious falsehoods. Once familiar with the context of the argument, we might correctly assess it as a woefully inept one, not redeemed by its veneer of validity. Similarly, the argument "Wong has been giving lavish presents to Ho, therefore, Wong is in debt" might be a very shrewd basis for the claim that Wong is in debt, despite its not being valid. Assessing the merit of an argument is not in general an a priori enquiry. And even when it is, say, in mathematics, that does not mean that 'we' can do it. Most of us would not present ourselves as judges of mathematical arguments, and even a distinguished mathematician is apt to be modest about checking some candidates.

The principal upshot here is this: one may have good reason to reject an argument (or a conclusion) without having assessed the argument and without even being able to do so if allowed to study it. It would be a fallacy to mistake this

reasonableness for having an objection to the argument. And this is a distinction which is difficult to articulate and thus a dangerous confusion which can be hard to avoid. The “*ad hominem* fallacy” is a classification which can serve us in warning against the confusion. It is worth reinforcing this classification.

One way is this: that P is selected by a procedure highly likely to pick falsehoods is good reason to reject P, but it is not at all *evidence* against P. To be evident is to be obvious, in plain view. Evidence for P is something which tends to make it evident that P or is offered as tending to do this. Adequate evidence for P succeeds in making it evident that P. Contemporary use of ‘Evidence’ is not always in accordance with this standard, but it would be good to return to it. Not all cases of giving good reasons for accepting or rejecting P are cases of giving evidentiary reasons.

Consider this case: You are a prisoner of a terrible tyrant, King Mog, who is known always to keep his promises. He tells you that tomorrow you will be given a lie detector test to determine whether you believe M: Mog is a kindly, enlightened monarch. If you do, you will be given \$10,000 and released unharmed. If you don't, you will be skinned alive and salted. Furthermore, a harmless, non-addictive drug will be offered to you, and if you take it while listening to a recording praising King Mog's kindness, you will wake up believing M. Mog argues to you as follows: (i) If you believe M, which you can easily do, you get a reward. (ii) If you don't believe M, then you will suffer horribly. (iii) Therefore, M. Has Mog given you good reasons for believing M? He has given you very good motivation for accepting M, but wretched evidence. I would mark this by saying that he has given good motivational reasons but bad evidentiary reasons. This is characteristic of fallacious appeals to emotion---in this case, the fallacy of *argumentum ad baculum* or appeal to force. Note that if the conclusion M were replaced with “You ought to accept M” then the reasons offered are good both motivationally and evidentially.

The fact that P was selected by a process likely to pick falsehoods is not a merely motivating reason as in the above appeal to force. But it is not evidence of the falsity of P either. Given the going loose use of ‘evidence’ it would be counted, but by a proper standard it should be otherwise named, say, as ‘probabilistic reasons.’ Just as the *argumentum ad baculum* has (typically, when fallacious) motivating reasons posing as evidence, the *argumentum ad hominem*, in its most interesting and formidable form, typically has probabilistic reasons posing as evidence.

Of course a good reasoner may find good evidentiary reasons to be good motivating reasons and also good probabilistic reasons. And just as moving from P

to “You had better accept P” can make the reasons which were good motivation but bad evidence into adequate evidence, so too changing from P to “It is highly likely that P” can make reasons which were good probabilistically but bad evidence into adequate evidence.

Thus in pattern I, the fact that P has been advanced by someone unlikely to be right about such questions is good reason to doubt that P is true. It is just not evidence for the falsity of P. The *ad hominem* fallacy confuses these two, so that the considerations which make it unlikely that P are taken to be considerations against P. It is easier to see this mistake in the above cases because we do not even know what proposition P is. We just know it is some proposition picked by a method which is highly likely to pick falsehoods.

However, this can become blurred as we gradually get some understanding of P. “Knowing what proposition P is” is very vague. One knows it is the proposition whose number was drawn from the urn. Or one may know that it is a proposition about a certain person, to the effect that that person grows melons. Or one may know who that person is, etc. It is easy to get the illusion of understanding the proposition and having evidence it is false when in fact one has nothing but probabilistic considerations.

“To conclude that P” is vague, but one good reading is in terms of claiming to know on the basis of premises. To claim to know that P is inconsistent with allowing that there is a nonzero chance that not-P. If you conclude that someone's claim that P, is unjustified, then you are (in the relevant uses) claiming to know that his claim that P, is unjustified. This cannot be done merely on the basis of probability. If you admit that there is any nonzero chance that his claim is justified then you cannot consistently claim to know it is not.

If we understand ‘evidence’ as recommended above (and it must be stressed that this is not an appeal to correct ordinary usage but rather, a correction of ordinary usage) then we can clarify the fallaciousness of pattern I(i) by distinguishing further between the conclusion that (ia) ‘we have evidence showing that A has not produced argument sufficient to justify his claim that P’ and that (ib) ‘it is likely that A's claim is not justified’. Iia should be seen as a kind of *ad hominem* fallacy, the justification *ad hominem*.

This point can be extended. Not only is Iia fallacious. It is fallacious to conclude from the premises of I even that you have any evidence undermining the claim that P, even slightly. One might put this by saying that it is fallacious to think that considerations showing that it is unlikely that P constitute any kind of defeater for the claim that P, except that the term ‘defeater,’ which has in its natural use the proper suggestion of a counterargument or evidence to the

contrary, has been appropriated as a technical term which can be used on the basis of merely probabilistic considerations.

In this connection, one might be tempted to appeal to a distinction between asserting the negation of a proposition and asserting that the proposition is false. When you do not know what the proposition is and are unable to present it, you can still assert it is false (as in "What he said yesterday is false") but you cannot assert the negation (as in "It is false that his program will reduce unemployment"). However, this is not helpful. You may 'know' that the proposition was that Wong grows melons in Daigan and have excellent probabilistic reasons for asserting the negation, that Wong does not grow melons in Daigan while still being utterly lacking in evidence for or against what you are asserting. Whatever logical difference may be found between asserting that a proposition is false and asserting its negation, or epistemological difference, will not be of use here. The right guide is our rule that evidence that the source of a claim is likely to be wrong is not evidence against the claim. The tendency to overlook this is the essential feature of the *ad hominem* fallacy.

It is a prominent feature of Descartes' philosophy to hold that we cannot know anything about the external world unless we know that (G) we are created by a benevolent God who has given us senses which are trustworthy if used with caution we are capable of exercising. I believe that G is true. But it would be a bad mistake to think that someone who does not believe that G and thus does not know that G could therefore be shown not to know anything about the world. Even worse would be the conclusion that someone who regards his thinking as made possible by Godless arrangements of matter largely determined by chance is, in taking this attitude, advancing a hypothesis which undermines his theorizing about the world or himself. Theories to the effect that you are a creature of a kind very unlikely to produce a good theory have been advanced both on theistic and atheistic grounds. In any case, the fact that such a theory would assign a low probability to the claim that its author has produced a good theory is not, on the proper understanding of evidence, evidence against the theory.

It is true that evidence that someone is utterly unreliable can be a reason for denying that he knows something. Suppose that an extremely unstable alcoholic mathematical genius, Jones, proves a brilliant theorem, say, that (H) all hypersimple manifolds are superousian. We might well have good reason for the verdict that Jones does not know that H. Jones might frequently deny H, or defend H by appeal to certain irrelevant facts about patterns in which his empty bottles break against the wall. We may find it nearly incredible that Jones got himself together enough to produce the proof. This sort of thing is indeed relevant to the

question whether Jones knows that H. But it is a bad confusion to think it counts as any kind of objection to Jones' argument for H. That Jones thinks that H could be a terrible reason for H. To advance it as a reason would be an *argumentum ad verecundiam*, an argument we can correctly reject. But that should not be confused with evidence against H, or we will slip into a fallacious *ad hominem*.

We can compare Jones' case with that of the earlier reputed 'idiot' who produces a brilliant argument. In either case, that argument is evidence that the arguer is capable of cleverness, is not altogether incompetent. The presentation of the argument tends to make this evident. When the 'idiot' was arguing that he is mentally incompetent, his reasoning was failing to make it evident that he is incompetent. The display of reasoning was in fact making it evident he has at least some competence.

Our model of merely probabilistic reasons was based on a random drawing model of probability—an argument drawn from an urn containing 10,000 proven bad arguments and one untested one. The argument that the evolution of human argumentative competence was highly unlikely would not plausibly be based on such a model. This is a limitation on charges of *ad hominem* reasoning in accordance with our account of such reasoning. That is only due to limitations on judgments of probability. Whether, for example, the development of human reasoning ability in a naturalistic evolutionary process has some probability, low or otherwise, is far less clear than the simplified probabilities based on urn drawings. Hopefully such unclarity in the notion of probability does not undermine the point that, insofar as we have a distinction between probabilistic and evidentiary reasons, it is fallacious to confuse them.

CONCEPTS, INTUITIONS AND EPISTEMIC NORMS

Murray CLARKE

ABSTRACT: In this paper, I argue that Dual Process Theories of cognition, as recently defended by Keith Frankish and Jonathan Evans, Keith Stanovich, Peter Carruthers, Richard Samuels, and others, offer a useful framework that can transform our conception of the nature and role of concepts in cognitive science and the role of intuitions in epistemology. The result is that recent debates concerning competing accounts of concepts, the role of intuition in epistemology, and debates between internalists and externalists concerning the nature of epistemic justification and knowledge, can be usefully advanced given the resources of such Dual Process Theories.

KEYWORDS: dual process, intuition, concepts,
internalism, externalism.

Some would say that concepts give rise to epistemic intuitions, and epistemic intuitions ground epistemic norms. Suppose that is true. A bit of crafty skullduggery would be all that is necessary to provide an explanatory story adequate for the linkages and we would be off to Belmont. But not so fast! One would still need an account of concepts in order to enjoy the races. But providing such an account of concepts turns out to be a nontrivial task. Be that as it may, I favor an evolutionary version of informational atomism unlike the originator of informational atomism, Jerry Fodor. Still, further problems await since any account of concepts would seem to be parasitic on an account of the mind. Here I have gone on record as defending the massive-modularity account of Cosmides and Tooby over against the peripheral modularity view of Fodor. Lately, however, proponents of the dual process theory (such as Keith Stanovich, Keith Frankish and Jonathan Evans, Peter Carruthers, and Richard Samuels) have changed the game. In what follows, I will clarify some of the claims of dual process theorists in section one. In section two, I will evaluate the implications of dual process theory for the cognitive science debate surrounding the nature of concepts. In section three, I will evaluate the implications of dual process theory for debates concerning the nature and role of intuitions in epistemology as it relates to the

internalist/externalist debate concerning the nature of epistemic justification and knowledge.

1. Dual-Process Theories

Dual-Process theorists argue that there are two minds in each cranium. These two minds employ two distinct processing mechanisms and employ different procedures to deal with deductive reasoning, decision making, and social judgment. As Keith Frankish and Jonathan Evans put it: “Typically, one of the processes is characterized as fast, effortless, automatic, nonconscious, inflexible, heavily contextualized, and undemanding of working memory, and the other as slow, effortful, controlled, conscious, flexible, decontextualized, and demanding of working memory.”¹ And dual-process accounts of learning and memory have also been developed, “...typically positing a nonconscious implicit system, which is slow learning but fast access, and a conscious explicit one, which is fast learning but slow access.”² Human cognition is then seen as involving two multi-purpose reasoning systems, System 1 and System 2. The former have the fast characteristics and the latter, the slow characteristics. Of course, there are a variety of differences among the positions held in this debate. Jonathan Evans provides the following chart of typical properties of System 1 and System 2 theories of cognition:

System 1

Evolutionarily Old
Shared with animals
Unconscious, preconscious
Controlled, volitional
Fast, parallel
Associated with language
Associative
Belief-based, pragmatic reasoning
Implicit knowledge
Independent of cognitive capacity
Personal

System 2

Evolutionarily New
Distinctively human
Conscious
Automatic
Slow, sequential
Independent of language
Rule-based
Abstract, logical-reasoning
Explicit knowledge
Dependent on cognitive capacity
Subpersonal³

¹ Jonathan Evans and Keith Frankish, eds., *In Two Minds: Dual Processes and Beyond* (Oxford: Oxford University Press, 2009), 1.

² Evans and Frankish, *In Two Minds*, 1.

³ Evans and Frankish, *In Two Minds*, 34.

It is often claimed that System 1 is early evolving, shared with other animals and includes implicit learning and modular cognition. In contrast, System 2 is recent, uniquely human, and is related to working memory and general intelligence. As such, System 1 is much more like the Cosmides and Tooby massive-modularity position in the recent literature⁴. Whereas, Fodor's peripheral modularity combined with a nonmodular, general intelligence capacity at the center of the mind is more like the hybrid System 1/System 2 combination. For Evans, in contrast, the idea is that System 1 and System 2 are responsible for type 1 and type 2 processing. Type 1 processes are fast, automatic, have high-processing capacity and require only low effort, while Type 2 processes are slow, controlled, of limited capacity, and require high effort to utilize. Evans follows Wason and Evans⁵ in using the type terminology. Evans then adds a System 3 processor to deal with conflict and control issues concerning the interaction of System 1 and System 2. On his view, these systems are tokens. In contrast, Richard Samuels has recently argued that the token systems view is mistaken as he argues for the system type or cognitive kinds position. Samuels also makes the important point that each system involves clusters of co-varying properties. That is, processes that exhibit one property typically possess the other properties. This matters since the fact that clusters exist suggests an underlying suite of mechanisms subserving such co-variation. This abductive inference then paves the way to posit a "bipartite division between cognitive mechanisms."⁶ In short, there are natural kinds that underwrite cognition where Samuels construes 'natural kind' in Richard Boyd's sense as homeostatic property clusters. According to Samuels, dual-process theorists endorse two claims:

1. Dual-Cluster Thesis: cognitive processes tend to exhibit either the S1 or S2 property clusters.

⁴ See Leda Cosmides and John Tooby, "Origins of Domain-Specificity: The Evolution of Functional Organization," in *Mapping the Mind: Domain Specificity in Cognition and Culture*, eds. Lawrence A. Hirshfeld and Susan A. Gelman (New York: Cambridge University Press, 1994), 85-116.

⁵ P. C. Wason and J. St. B. T. Evans, "Dual Processes in Reasoning," *Cognition* 3 (1975): 141-54.

⁶ Richard Samuels, "The Magical Number Two, plus or minus : Dual-Process Theory as a Theory of Cognitive Kinds," in *In Two Minds: Dual Processes and Beyond*, eds. Jonathan Evans and Keith Frankish (Oxford: Oxford University Press, 2009), 131.

2. Dual-Systems Thesis: there is a division in our cognitive architecture—a division between cognitive systems—that explains this clustering effect.⁷

Samuels thinks that there are two ways of developing these generic claims: these are the token and the type thesis. The Token Thesis maintains that there are just two particular cognitive mechanisms or systems. The System 1 mechanism subserves cognitive processes that exhibit the S1-property cluster. The System 2 mechanism subserves cognitive processes that exhibit the S2-property cluster. Each human mind exhibits a fundamental, bipartite division into these two particular systems. Evans accepts a version of this claim because he argues that we have an old mind and a new mind that consists, respectively, of a group of System 1 and a group of System 2 processes. In contrast, according to the Type Thesis, each mind is constructed out of two types or kinds of cognitive system. Systems of the first kind subserve processes that tend to exhibit the S1-cluster. Systems of the second kind subserve processes that tend to exhibit the S2 cluster. But there is no overarching old mind/new mind dichotomy or Token Thesis at play. The Token Thesis implies the Type Thesis but not vice-versa. As such, the Type Thesis is logically weaker than the Token Thesis. Samuels defends the Type Thesis as more plausible than the Token Thesis principally because there seem to be *many* system 1 and system 2 devices in the mind. For instance, the human visual system involves many subsystems for depth perception, color identification, and categorization.⁸ And, these subsystems themselves decompose into smaller units, and so on. Now one might try to avoid the trivialization of the token thesis by relativizing the claim to some, fairly abstract, level of decomposition. But, as Samuels says,

...even at quite abstract levels of decomposition, it's just not plausible that our minds contain only two systems. On any plausible decomposition, there are likely to be a great many systems for a wide range of different mental processes, including perception, memory, reasoning, emotion, language, and no doubt many others. Moreover, it's not plausible to treat all these devices as constituting just two systems. Not, at any rate, unless one is prepared to countenance systems that are wildly heterogeneous in character.⁹

One might like to defend the idea that there are exactly two reasoning systems in each mind. Samuels demonstrates that this view cannot be sustained

⁷ Samuels, "The Magical Number Two," 132.

⁸ Stephen E. Palmer, *Vision Science: Photons to Phenomenology* (Cambridge: MIT Press, 1999).

⁹ Samuels, "The Magical Number Two," 134.

under a variety of construals of the key notion of what counts as a reasoning system. Now dual-process theorists acknowledge Samuels point here but often claim that the collapse of the Token Thesis is not two-way. That is, there are many System 1 mechanisms but only one System 2 mechanism. But Samuels finds this just one System 2 mechanism claim dubious too since the processes and mechanisms posited in each System 2 domain are not characterized in the same way. For instance, in social cognition, Matthew Lieberman and colleagues posit a mechanism for controlled social cognition, the C-System, whose processes exhibit many S2 properties but these properties vary from the properties that researchers posit who work on deductive reasoning.¹⁰ Maybe researchers are moving toward a single System 2 for reasoning across all these domains but Samuels doubts that this goal will be achieved, preferring the idea that the researchers are identifying different mechanisms that subserve processes of the same general type, i.e., the S2 cluster. Suppose that Samuels is correct and the cognitive kinds or Type Thesis is correct, how would this bear on issues concerning concepts, intuitions, and epistemic norms? I propose to discuss these issues in two separate sections. In section two, I will discuss concepts and, in section three, I will discuss intuitions and epistemic norms.

2. Concepts

Nativists, like Fodor, defend the idea that all of our lexical concepts are not learned while more moderate nativists, such as Susan Carey in her recent book, *The Origin of Concepts*, maintain that only core cognition principles are innate.¹¹ In *The Language of Thought*, Fodor argued that all of our lexical concepts are not learned because they are all innate.¹² But critics thought it somewhat unlikely that ‘carburetor’ and the other 500,000 primitive lexical concepts are innate. In her recent live stream debate with Jay McClelland at Ohio State University, Carey called radical concept nativism ‘absurd.’¹³ Fodor’s response to this sort of criticism was contained in his book, *Concepts*, where he rescinds radical concept nativism by arguing that our primitive lexical concepts are neither learned nor innate, but

¹⁰ Matthew Lieberman, “Reflective and Reflexive Judgement Processes: a social cognitive neuroscience approach,” in Joseph P. Forgas, Kipling D. Williams, and William von Hippel (eds.) *Social Judgements: Implicit and Explicit Processes* (Cambridge: Cambridge University Press, 2010), 44–67.

¹¹ Susan Carey, *The Origin of Concepts* (New York: Oxford University Press, 2009).

¹² Jerry Fodor, *The Language of Thought* (Cambridge: The MIT Press, 1975).

¹³ Susan Carey’s debate with Jay McClelland at Ohio State University, April 16, 2010.

acquired.¹⁴ On his informational atomist view, we ‘lock to’ or ‘resonate to’ mind dependent concepts by virtue of innate neurological mechanisms. First, we learn the doorknob stereotype by experiencing good instances of doorknobs and then we acquire the concept doorknob by virtue of innate neurological mechanisms. The first step is psychological; the second, biological. In contrast, Stephen Laurence and Eric Margolis defend the learning part of Fodor’s account but reject the biological part and so defend a straight learning account of concepts.¹⁵ Fodor’s updated two-step account in *The Language of Thought Revisited* attempts to retell the locking story by appeal to an attractor landscape metaphor but represents no change, or improvement, in his view.¹⁶ So there is an impasse here between Laurence and Margolis on the one hand, and Fodor, on the other. This impasse was mirrored in the recent live stream debate between McClelland and Carey at Ohio State University where Carey played Fodor against McClelland’s Laurence and Margolis in the broad sense that Carey is a nativist about concepts and human knowledge while McClelland defended a learning theory about knowledge and concepts. Of course, Carey’s nativism is much more moderate than Fodor’s, and McClelland’s learning theory is a connectionist associationism which Laurence and Margolis have never defended.

I mention these debates between nativists and learning theory folk about the status of concepts and knowledge for one reason. Dual-process theory may provide a route out of multiple impasses that exist in the cognitive science literature. What I have in mind is that we may not need to choose between a learning account of concepts (and knowledge) and a locking, acquisition account of concepts. Maybe the concepts that are ‘locked to’ are the one’s involved in System 1 processes, while the concepts that are learned are the product of System 2 processes. This would provide an interesting resolution to a debate that has raged for a very long time: some concepts are acquired and some are learned. Now I don’t think my suggestion will come as any surprise to many. After all, Carey in her live stream “Origins of Knowledge” talk in April said (approximately) that: “The key issue is not if there are any innate perceptual primitives but how much is innate.”¹⁷ The key issue is the extent of the innate conceptual and knowledge mechanisms in place. Carey thinks that the face recognition mechanism, inter alia,

¹⁴ Jerry Fodor, *Concepts: Where Cognitive Science Went Wrong* (New York: Oxford University Press, 1998).

¹⁵ Eric Margolis and Stephen Laurence, eds., *Concepts: Core Readings* (Cambridge: The MIT Press, 1999).

¹⁶ Jerry Fodor, *The Language of Thought Revisited* (New York: Oxford University Press, 2008).

¹⁷ Susan Carey in her live stream “Origins of Knowledge” Talk in April 2010

is a clear example of such an innate structure. Of course, she accepts that once we have such structures then McClelland, and others', connectionist models become very important for clarifying how learned concepts emerge. Hence, any adequate acquisition story must show how innate conceptual structures and mechanisms give rise, with the aid of informational input to learned concepts and knowledge via, perhaps, connectionist models. Carey thinks that core cognition is not core knowledge, as Elizabeth Spelke calls it¹⁸, because such representations need not be veridical. Core cognition is shared with nonhuman animals and is the developmental foundation for human conceptual understanding. As she notes: "Like sensory and perceptual features of the world, the entities in core domains of knowledge are identified by modular innate perceptual-input devices. Therefore, the extension of the symbols that articulate core cognition is fixed in part, by evolutionarily underwritten causal relations between entities in the world and representations in the mind."¹⁹ Such core cognition representations differ from fully explicit conceptual representations that pick out intuitive theories. Causal connections mediated by perceptual-input analyzers determine the referents of core cognition, while the explicit conceptual representations of intuitive theories also involve social processes via inferential roles for their articulation (ala Kripke, Putnam, and Burge). On Carey's view, only humans create conceptual representations that go beyond sensory representations and core cognition. Humans create new representational resources that are 'qualitatively different' from the representations that they are built from. As such, she denies the Fodor/Macnamara continuity thesis that all the representational and inferential capacities that underlie adult belief systems are present throughout development or arise from processes such as maturation. Fodor calls a version of this thesis connected to language the compositionality constraint on concepts. For Fodor, since concepts are the constituents of thoughts and often of each other then mental representations inherit their contents from the contents of their constituents. Hence, complex concepts decompose into primitive concepts and their relations. The result is the familiar Fodorian claim that belief is productive and systematic. Productive, since there are an infinite number of beliefs that one can entertain and systematic, since the ability to entertain one thought implies the ability to entertain lots of others related to the content of the first one.²⁰ Carey

¹⁸ See, for example, Elizabeth S. Spelke and Katherine D. Kinzler, "Core Knowledge," *Developmental Science* 10, 1 (2007): 89–96.

¹⁹ Carey, *Origins*, 11.

²⁰ Fodor, *Concepts*, 26.

denies this claim insisting that conceptual development is discontinuous. Humans are capable of novel thoughts using novel concepts. For example, the integer list is a cultural construction with more representational power than any of the core representational systems on which it is built. The notion of a rational number, likewise, transcends the integer representations available from the outset of the construction process in ontogenesis. Children, in effect, create incommensurable new concepts through maturation. Kuhnian conceptual revolutions occur due to Quinian bootstrapping mechanisms, Carey thinks, like those discussed in the history and philosophy of science by Nancy Nersessian.²¹ Theoretical knowledge that transcends core cognition is facilitated, by such, Quinian bootstrapping. She also notes that such intuitive theories are not innate, the entities in their domain are not identified by input analyzers, their format is not iconic, and they are not continuous throughout development.²² Carey has the System 1/System 2 distinction in mind here in the sense that she accepts that the concepts of core cognition are evolutionarily old, result from modular processors that are fast, shared with animals, associative, involve implicit knowledge, and are associated with language. In contrast, the intuitive concepts of explicit theory are evolutionarily recent, uniquely human, conscious, slow and sequential in production, involve abstract, logical reasoning, explicit knowledge, and so on. As such, she endorses the idea of distinct kinds of concepts and knowledge for each system. It should also be noted that some representations in core cognition may be nonconceptual as with early perceptual representations.

My own view is that informational atomism combined with an evolutionary acquisition account best captures the concepts of System 1, while definitionism or the classical theory of concepts captures one example of a System 2 learning account of concepts. Of course, there are other examples of System 1 and System 2 concept accounts on offer. Fodor and Carey provide two distinct examples of how one might provide distinct accounts of concepts that depend on the System 1/System 2 distinction. My examples are simply meant to illustrate what can count as an account of concepts relative to System 1 or System 2. The point that I want to drive home is that we need a distinct account of concepts for each System. This is a crucial point and one, in effect, denied by, for instance, Alison Gopnik and

²¹ Nancy Nersessian, "How do scientists think? Capturing the dynamics of conceptual change in science," in *Cognitive models of science*, ed. Ronald N. Giere (Minneapolis: University of Minnesota Press, 1992), 3-44.

²² Carey, *Origin*, 22.

Andrew N. Meltzoff²³ when they claim that scientists and young children share the same type of representations when they theorize. On my view, and Carey's view, this cannot be so because distinct types of representations attach to children's core cognitive representations as opposed to their mature intuitive theories that are developed once they become adults. The child has no theories in the sense that the scientist does precisely because children fail to have the same kind of concepts that the scientist has. In effect, children have no intuitive theories at all. With this understanding of the distinction between two kinds of concept in mind, I now want to see what the implications might be for our understanding of intuitions, and the role of intuition, in philosophical theorizing.

3. Intuitions and Epistemic Norms

Alvin Goldman has argued that a-priori intuitions are the product of our concepts.²⁴ As Hilary Kornblith notes concerning Goldman's view about the relation between concepts and intuitions:

Armchair methods in philosophy work as well as they do, according to Goldman, because there is a certain kind of causal relationship between our concepts and our intuitions. In particular, our concepts are causally responsible for our intuitions; more than this, the manner in which our concepts bring about our intuitions makes our intuitions reliable indicators of the truth of their contents. So if intuitions and concepts are related in this sort of way, the armchair methods employed by philosophers will be extremely revealing of the nature of our concepts.²⁵

Suppose Goldman is correct. We might still ask whose intuitions we should appeal to. What are pre-theoretic intuitions? As Kornblith points out, Frank Jackson thinks conceptual analysis involves characterizing widely shared 'folk concepts.' This would seemingly have us consulting large groups of people about their intuitions. In fact, both Goldman and Jackson advocate soliciting pre-theoretic, spontaneous, intuitions from one's students. Philosophical practice, in contrast, has relied heavily on the intuitions of professional philosophers whose

²³ Alison Gopnik and Andrew N. Meltzoff, *Words, Thoughts, and Theories* (Cambridge: MIT Press, 1998).

²⁴ Alvin I. Goldman, "Psychology and Philosophical Analysis," and "Epistemic Folkways and Scientific Epistemology" in his *Liaisons: Philosophy Meets the Cognitive and Social Sciences* (Cambridge: MIT Press, 1992), 143-153, 155-175.

²⁵ Hilary Kornblith, "Naturalism and Intuitions," in *Philosophical Knowledge: its Possibility and Scope*, eds. Christian Beyer and Alex Burri (Amsterdam-New York: Rodopi, 2007), 30.

intuitions have been shaped by years of study and the shaping of intuition by epistemic concepts like justification and knowledge. These reflective intuitions are anything but the pre-theoretic intuitions of the folk. Jackson suggests that we should consult the intuitions of our students but these intuitions are not pre-theoretic either. Typically, such intuitions are only sought once a great deal of background theory has been force fed to the supplicants. And, as Kornblith points out, the public showing of hands is subject to a number of biases, such as the anchoring effect.²⁶ Kornblith also points out that: "Finally, the very sort of controls which psychologists routinely bring to bear on experiments of this sort, such as controlling for order of presentation, are rarely if ever brought to bear on philosophy classroom surveys."²⁷ This suggests that philosophical practice must be dramatically changed if the goal is to elicit pre-theoretic folk intuitions. On the other hand, Kornblith thinks that it is odd that we should consult folk intuition at all when doing naturalized epistemology, we would never do that in science. Why would we do that in philosophy? And, where conceptual analysis has had some successes, we should not simply alter what we do. It is a truism in philosophy of science that observation is theory infected. The goal, therefore, is simply to infect observation with the correct theory.²⁸ As Kornblith notes, the idea that observation might obtain without dependence on theory at all is now simply taken to be a logical positivist ideal that few would want to defend today. This is an interesting point that Kornblith makes but keep in mind that Carey thinks that not all observation is theory-laden. In particular, core cognition and perceptual primitives do not constitute theories at all. In effect, Carey rejects Quine's claim that the simple observation that 'that is an object' is the result of a theory. Intuitive theories come much later in development. Hence, even if theory does infect some observation, it does not infect all observation. This is a crucial point. Moreover, the intuitions that are the product of System 1 would be distinct from the intuitions that are the product of System 2. In effect, philosophers of science have been guilty of thinking that there are only observations that are framed by System 2 concepts and their intuitions. This is false. There are observations framed by concepts from System 1 and their intuitions. The well-worn Muller-Lyer Illusion is one such example where our observations are not informed by

²⁶ Kornblith, "Naturalism," 33.

²⁷ Kornblith, "Naturalism," 33.

²⁸ For my purposes, I am going to make the neo-positivist assumption here that theories are sets of sentences that decompose into concepts in various syntactic relations. In an extended sense, I will also assume that theoretical concepts give rise to theoretical intuitions.

background theory so that we continue to see lines of the same length as unequal due to the modularity of our perceptual apparatus and the concepts and intuitions that inform this module.

If this point generalized which it no doubt does, then we might expect that intuitions that are the product of System 1 concepts would produce fast, pre-theoretic, raw intuitions. In contrast, the more careful, slow, reflective intuitions characteristic of philosophy would be the result of System 2 processes and concepts. This might help to make sense of the recent literature on intuition in epistemology. Instead of a zero sum game where competing accounts of intuition are defended, we could simply recognize that some intuitions have their source and role in System 1 concepts and other intuitions have their source and role in System 2 concepts. This would constitute an interesting result that would help philosophers adjudicate between competing accounts of intuition. The untutored intuitions of the unwashed nonphilosophical public concerning knowledge would simply lead to our knee-jerk understanding of the acquired, System 1, concept of knowledge common to all species (and defended by externalists and reliabilists) while the sophisticated intuitions of the philosopher would be employed to render the reflective notion of knowledge unique to humans and sought after by epistemic internalists. Now some naturalists, such as Kornblith and Clarke, will not be happy with this result because they think that knowledge is not a concept at all and that the phenomena of knowledge is a univocal, natural kind (Kornblith) or set of natural kinds (Clarke) to be discovered in the world. How might we respond to these worries here? We might begin by suggesting that, at the very least, since there are two kinds of concepts related to each of System 1 and System 2 processors that there are two kinds of intuitions parasitic on these two kinds of concepts. Now if that is true, then when internalists refer to knowledge then they must be referring to the System 2 reflective concept of knowledge and the intuitions that underlie it. If that is right, then it would seem to follow that perceptual knowledge from the internalist standpoint is a misnomer. Perceptual knowledge must be different in kind from reflective knowledge since the intuitions that support it must be different in kind from the intuitions of System 2 knowledge. That is, perceptual knowledge must be the result of System 1 concepts and intuitions. In contrast, reflective knowledge must be the result of System 2 concepts and intuitions. For an internalist, knowledge and justification must turn out to be bipartite according to their own lights. Nonreflective knowledge, if this is correct, has been largely ignored in the internalist tradition in the sense that internalists have tried to explain nonreflective knowledge in reflective terms, not understanding that a completely different analysis of it was necessary. One thinks

here of the somewhat turgid and unlikely account of perceptual knowledge of C.I. Lewis²⁹. But what about externalists, like Kornblith: what epistemic sins has he committed? Kornblith (following Goldman, Dretske and other externalists) denies that knowledge requires any sort of reflection, contra internalists. If that is so, he has ignored the concept of reflective knowledge involved with System 2 mechanisms and the intuitions that underlie that type of system. Kornblith would say that the concept of nonreflective knowledge and the concept of reflective knowledge are one and well worth studying. However, his interest is in the phenomena of knowledge in the world. The study of such a natural kind, like any other natural kind, requires that we study the empirical theory of, for instance, ethologists and see how they refer to the knowledge of, say, the piping plover. In this way, we will begin to understand the phenomena of knowledge as it occurs in the wild where empirical theory is squared with the empirical judgements we make when we see birds, for instance, protecting their young from predators by engaging in broken wing displays and so on. Kornblith suggests that we should take the language of ethologists literally when they talk about piping plover knowledge because that talk plays an essential role in successful, empirically informed, theory. Theory is squared with informed empirical judgement on his view, not with nonreflective or reflective intuition. But this is not a problem since we are no longer attempting to *square* our *concept* of knowledge with *intuitions* about knowledge. Of course, an interesting residual question here concerns the relation that might obtain between our *concepts* of knowledge and the *phenomena* of knowledge. Kornblith glides over this important issue in his 2002 book, *Knowledge and Its Place in Nature*.³⁰

Another issue concerns the relation between theory and evidence as opposed to concepts and intuitions. The philosophical literature, as Jennifer Nagel points out, contains two conceptions of the relation between “particular case intuitions and more general theories in epistemology.”³¹ She notes, following Stich in *The Fragmentation of Reason*,³² that advocates of reflective equilibrium, such as Nelson Goodman, maintain that philosophical progress is made by adjusting general theories to better match our judgments about particular cases while also adjusting our judgments about particular cases to conform to our general

²⁹ From Clarence Irving Lewis, *An Analysis of Knowledge and Valuation* (LaSalle: Open Court, 1946).

³⁰ Hilary Kornblith, *Knowledge and its Place in Nature* (Oxford: Oxford University Press, 2002).

³¹ Jennifer Nagel, “Epistemic Intuitions”, *Philosophical Compass* 2, 6 (2007): 792-819.

³² Stephen Stich, *The Fragmentation of Reason: Preface to a Pragmatic Theory of Cognitive Evaluation* (Cambridge: The MIT Press, 1993).

theories.³³ Where Stich accuses Goodman of advocating a conservative policy here, Nagel notes that Goodman also allows for theoretic reform where, due to convenience or theoretical utility, we allow a theory to run counter to the mandates of common usage. As such the new definition alters, rather than merely extends, current usage. In contrast, Rudolph Carnap³⁴ adopts a more radical conception where philosophical progress occurs by virtue of a process he calls ‘explication.’ On this view, we refine a messy and vague pre-scientific concept (the explicandum) into a simpler and more exact scientific term (the explicatum). As she notes: “A successful explication delivers an exactly defined term that applies to most of the terms once picked out by the explicandum; this new term should be both simple and fruitful, readily connected to an existing network of scientific concepts and helpful in the formulation of new laws.”³⁵ She also notes that where reflective equilibrium gives equal weight to particular and general judgments, explications assigns them very different roles. In explication, we start from particular cases but once the scientific definition is formed “it is not subject to further constraint from reflection on the intuitiveness of its application to particular cases.”³⁶ Instead, once we have the new scientific meaning of a concept our intuitions about cases are simply guided by the definition. Alternatively, we might use the definition where precision is required saving the ordinary explicandum for everyday use. Notice that determining folk epistemic usage ala System 1 might be facilitated by the Goodmanian reflective equilibrium method, while the method of standard analytic epistemology over the last fifty years seems more attuned to the Carnapian explication method ala System 2. Kornblith’s own naturalized version of studying the phenomena of knowledge as scientists use it, in contrast, is much more like adopting the results of the scientists as they use a scientific version of the Carnapian method of explication with one alteration. Kornblith denies that the appeal to particular cases, or intuitions, plays much role in science nor should it in philosophy. Such intuitions should be shelved as soon as observations become available for the construction of theory. In essence, philosophical theory construction should become scientific theory construction since philosophical theories should be empirical theories. Little wonder then that

³³ For Goodman’s defense of reflective equilibrium see his *Fact, Fiction, and Forecast* (Indianapolis: Bobbs-Merrill, 1965), 66-67.

³⁴ In Rudolph Carnap, *Logical Foundations of Probability*, 2nd ed. (Chicago: University of Chicago Press, 1962). See also his “P. F. Strawson on Linguistic Naturalism,” in *The Philosophy of Rudolf Carnap*, ed. Paul Arthur Schilpp (LaSalle: Open Court, 1963), 933-940.

³⁵ Nagel, “Epistemic,” 795.

³⁶ Nagel, “Epistemic,” 795.

the Carnapian notion of explication should sound a lot like Kornblith's proposed method: Carnap was also adapting scientific method for philosophical purposes. From such a perspective, philosophical method should simply be the scientific method. Scientific method, on the other hand, is the paradigmatic case of a System 2 processor. Science involves a slow, conscious, evolutionarily new, distinctively human, sequential, attempt at acquiring explicit knowledge by means of abstract, logical-reasoning given empirical inputs. But note also that we do not have to choose between the Goodmanian reflective equilibrium method and the Carnapian explication method of philosophizing. If our goal is to understand folk epistemology, then we should use the Goodmanian method since it operates using System 1 intuitions alongside System 2 reflective processes. In contrast, if our goal is to understand the exact concept of knowledge or the phenomena of natural kind or kinds that are knowledge, the Carnapian notion of explication is what is needed and the system employed, essentially, will be System 2. Of course, this latter natural kind inquiry will have as output, not the concept of knowledge, but a well-developed theory of the natural kind or kinds that constitute the phenomena of knowledge.

Despite Kornblith's claim from *Knowledge and its Place in Nature* that the phenomena of knowledge is the same in animals and humans and that no reflection is essential for knowledge, he has recently added that reflection, where it occurs, is similar in humans and nonhuman animals. In "The Myth of Epistemic Agency," his April Invited Lecture at Northwestern and also in *Knowledge and its Place in Nature*, Kornblith argued that Sosa's reflective knowledge versus animal knowledge distinction, cannot withstand scrutiny.³⁷ Why? The problem is that the divide implicit between animals and humans concerning reflection is mistaken. The commonsensical picture that human knowledge involves reflection while animal knowledge does not is just false because it underestimates the sophistication of animal knowledge and over intellectualizes human knowledge. Both humans and animals update their beliefs about the whereabouts of objects without any reflection. As he notes:

If whenever I see a fox approach, I come to believe that it is dangerous, the discovery that a particular fox is harmless will not be something that I simply register atomistically; it will bring about a change in the inferences I draw when I

³⁷ Hilary Kornblith, "The Myth of Epistemic Agency" (Graduate Philosophy Conference, Invited Address, Northwestern University, 2010). For Sosa's reflective knowledge versus animal knowledge distinction see Ernest Sosa, *A Virtue Epistemology: Apt Belief and Reflective Knowledge, Volume 1* (Oxford: Oxford University Press, 2007), 30-36.

am confronted with this particular fox-assuming, of course, that I can recognize it when I see it again. But the same sort of inferential integration, and change in inferential tendencies, can be found in many non-human animals, including for example, piping plovers. One needn't have anything like the cognitive sophistication of a primate, let alone a human being, in order to integrate information in this sort of way. The suggestion that this ability is a by-product of the ability to reflect, and thus, unique to human beings, is mistaken.³⁸

Humans, Kornblith thinks, do not utilize reflection in such cases and nonhuman animals integrate information in the same way that humans do. It is important to note that Kornblith accepts the point that human reflection does occur and that it is unique to us, it is just that we should not exaggerate its prevalence or reliability. He claims that both first-order processes and second-order reflective processes that result in belief are a mixed bag, sometimes they are reliable and sometimes they are not. In particular, some second-order reflective belief processes give rise to false belief. Certainly the catalog of belief biases in the cognitive science literature stands as a testimony to his claims here. Even once subjects are asked to reflect on their beliefs, they often become even more convinced that their false beliefs are true as Kornblith pointed out in Chapter Four of *Knowledge and its Place in Nature* concerning the position effect³⁹ and the anchoring effect.⁴⁰ In fact, reflection may reduce the accuracy of our beliefs and create the illusion of providing a real check on the reliability of our first-order, or System 1, processes. The other familiar Fodorian fact that Kornblith emphasizes is that because many cognitive processes are informationally encapsulated in cognitive modules, they will not be available to introspection and even if they were available, they will not be alterable by appeal to reflection. The by now clique example of this is the Muller-Lyer Illusion where background knowledge that two lines are the same length will not penetrate the perceptual module that determines that the lines are not the same length, where arrows appended in different directions at the end of each line lead us astray. The perceptual module, in other words, is informationally encapsulated from, or sealed off from, background information in such cases. For all of these reasons, the role and utility of reflection in human knowledge has been greatly exaggerated.

At any rate, this dual picture of concepts and intuitions has the potential to help us rewrite the recent history of epistemology in a way that is informed by the

³⁸ Kornblith, "The Myth."

³⁹ Kornblith, *Knowledge and its Place*, 111.

⁴⁰ Kornblith, *Knowledge and its Place*, 113.

results of cognitive science. But, perhaps, more importantly, it would help us determine which epistemic intuitions should be consulted when our goal is to understand epistemic norms.

4. Conclusion

Now consider: If epistemic norms are grounded in epistemic intuitions, and epistemic intuitions emerge from our concepts then if those concepts are parasitic on either System 1 or System 2 mechanisms it would seem to follow that knowledge and justification are the products of one, or both, of these two systems. In particular, if I am even close to the mark, the debate between externalists and internalists is really a debate that is parasitic on, and confuses, the contributions of System 1 and System 2 mechanisms. We would not have to choose between internalism and externalism in epistemology, the insights of both camps could be recognized. In this way, we would have dissolved some hotly contested debates in cognitive science and epistemology by recognizing that there was something right about the apparently incompatible positions involved in these debates. Resolution would be effected by dissolution.

The dual-process account allows us to relativize the insights of competing views in order to see the unique contributions of apparently competing authors more clearly. Of course, this sorting out of internalism and externalism only works if the naturalists/externalists, e.g., Goldman and Dretske, in question accept the notion that they are studying the concept of knowledge. For a naturalist/externalist who denies the claim that epistemology involves the study of our concepts of justification and knowledge, such as Kornblith and Clarke, things get more complex. If epistemologists should be studying the phenomena of knowledge in the natural order to discover the natural kind (Kornblith) or set of natural kinds (Clarke) that constitute knowledge then philosophical method must be dramatically changed to reflect such facts. For instance, one might study System 1 mechanisms and their inputs in order to understand the phenomena of System 1 knowledge. Equally, one would need to study System 2 mechanisms and their inputs in order to understand the phenomena of System 2 knowledge. My attempt in *Reconstructing Reason and Representation*⁴¹ to provide a modular account of knowledge can now be seen as an attempt to explicate the System 1 phenomena of nonreflective knowledge using System 2 reflective reasoning.⁴²

⁴¹ Murray Clarke, *Reconstructing Reason and Representation*. (Cambridge: The MIT Press, 2004).

⁴² My student, Guillaume Beaulac, made this point in conversation.

What now needs to be undertaken is the explication of System 2 reflective knowledge using the resources of System 2 reasoning. Subsequently, one would need to explain the relationship between our concept of reflective knowledge and the phenomena of knowledge. The sort of inquiry launched would radically transform the standard methodology of epistemology. In effect, epistemologists would become experimental philosophers. Of course, I have not attempted to clarify the details of the emerging nonreflective and reflective concepts of knowledge and the relationship between these concepts of knowledge and the phenomena of knowledge. Clearly, much work needs to be done on these issues. Still, the idea of integrating the new dual-process account of the mind with a revised epistemology is enticing because it promises rich philosophical rewards.

FREGE'S CONTEXT PRINCIPLE: ITS ROLE AND INTERPRETATION¹

Sorin COSTREIE

ABSTRACT: The paper focuses on Gottlob Frege's so called *Context Principle* (CP hereafter), which counts as one of the most controversial points of his philosophy. Due to its importance and centrality in Frege's thought, a detailed discussion of the principle requires a detailed analysis of almost all aspects of his philosophy. Obviously, such a task cannot be successfully accomplished here. Thus I limit myself to address only two questions concerning the CP: *what role does the principle play* (in *Grundlagen*) and *how can we interpret it*. Addressing the first problem is required in order to address the second. Most authors interpreted CP from the perspective of Frege's later distinction between sense and reference, which I will call the 'semantic interpretation'. Although I accept this perspective as valuable and important, I will initially inverse the action and I will try to approach CP, and generally *Grundlagen*, in a more natural way, contextually, namely setting them in the initial logicist plan of the *Begriffsschrift*. Finally, I will try to provide an interpretation concerning the alleged conflict between CP and Frege's compositionality thesis such that they could coherently stay together.

KEYWORDS: context principle, compositionality, sense,
reference

1. The Role of the Context Principle in *Grundlagen*.

1.1. Frege's *unity of thought*

There is development in Frege's thought, but seldom retractation, and, when does occur, it is usually in the nature of an emendation requiring little adjustment in the remainder of the system. This almost linear character of the development of

¹ This paper was made within *The Knowledge Based Society Project* supported by the Sectorial Operational Program Human Resources Development (SOP HRD), financed by the European Social Fund, and by the Romanian Government under the contract no. POSDRU ID 56815.

Frege's philosophy justifies the method (...) of *considering Frege's philosophy as a whole*, rather than as it existed at any particular stage.²

I am sympathetic with this view, and in fact this perspective provides me the *reading key* which entitles me to move conceptually back and forth from *Grundlagen*³ to both *Begriffsschrift*⁴ and *Grundgesetze*⁵, plus to any other later writings. I shall give three points in support of this view, especially with regard to the persistence of Frege's adherence to CP:

First, there is a clear *continuity* of problems through all his major works (the reduction of mathematics to logic, the rejection of psychologism and formalism, the logical power of his 'conceptual notation,'...etc); this issue will better clarify when I will discuss the connection between *Begriffsschrift* the three principles of *Grundlagen*.

Second, the main difficulty in claiming that the unity of Frege's thought was the *apparent* impossibility to accommodate in one coherent picture CP with Frege's later thesis regarding the compositionality of meaning. But, as I will try to show at the end of the paper, this alleged incompatibility can be dismissed and so the coherence of the system could be successfully saved.

Third, we should not forget Frege's *intellectual honesty*, and thus, since CP plays a central and explicit role in *Grundlagen*, an eventual rejection of it in later works would not have been passed tacitly, but surely it would have been signaled by an explicit statement, exactly like in the case when he acknowledged the catastrophic consequences of Russell's paradox for his theory.

² Michael Dummett, *Frege. Philosophy of Language*, second edition (Cambridge: Harvard University Press, 1981), 628; my italics in the original text.

³ Gottlob Frege, *Die Grundlagen der Arithmetik, eine logisch-mathematische Untersuchung über den Begriff der Zahl* (Breslau: W. Koenner, 1884) translated as Gottlob Frege, *The Foundations of Arithmetic*, trans. J.L. Austin, 2nd ed. (Oxford: Blackwell, 1953).

⁴ Gottlob Frege, *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens* (Halle: I. Nebert, 1879), translated in Gottlob Frege, *Conceptual Notations and Related Articles*, trans. and ed. Terrell Ward Bynum (Oxford: Oxford University Press, 1972), and selections in *The Frege Reader*, ed. Michael Beaney (Oxford: Blackwell, 1997).

⁵ Gottlob Frege, *Grundgesetze der Arithmetik*, 2 vols. (Hildesheim: Olms, 1962); preface, introduction and sections 1-52 of vol. I and appendix to vol. II translated in Gottlob Frege, *The Basic Laws of Arithmetic: Exposition of the System*, ed. Montgomery Furth (Los Angeles: University of California Press, 1964); parts of vol. II in *The Frege Reader*, ed. Michael Beaney (Oxford: Blackwell, 1997).

1.2. The continuity revealed in the case of *Begriffsschrift* and *Grundlagen*

Since Frege in *Grundlagen* is casting a great role for his three fundamental principles, one may rightly ask why he did not provide anything here to support them, in order to convince us why should we accept them so unconditionally⁶. One adequate answer would be that the problems addressed in *Grundlagen* arise directly from *Begriffsschrift* and thus it would be somehow superfluous to restate extensively all the guiding principles. But the credibility of such an answer lies on the detection of the principles in *Begriffsschrift*; therefore, they should be in *Grundlagen* only echoes of what has been already stated previously in there. The three *fundamental principles*, as they appear in the introduction of *Grundlagen*, are:

P1: Always to separate sharply the psychological from the logical, the subjective from the objective;

P2: Never to ask for the meaning of a word in isolation, but only in the context of a proposition;

P3: Never to lose sight of the distinction between concept and object.

But how are they related to previous points of *Begriffsschrift*? **P1** surely directs us to the idea of a 'pure thought,' which is central in *Begriffsschrift*, and which is secured by expelling any psychological ingredient out from our logic. **P2**, following Frege's own characterization of the principle ("if the second principle it is not observed, one is almost forced to take as meanings of words mental pictures or acts of the individual mind, and so to offend against the first principle as well"⁷), could be thus seen⁸ as a corollary of **P1**. **P3** is merely a reformulation of the technical and fundamental distinction between function and argument,

⁶ After stating them, Frege is mentioning very briefly some consequences for the system if they would lack; all of them are connected with his explicit and constant rejection of *psychologism* from both logic and mathematics.

⁷ Gottlob Frege, *Die Grundlagen*, x.

⁸ As we will see very shortly in detail, P2 has Kantian roots and thus could be also regarded as an elaboration of the 'priority thesis: the meaning of a sentence is prior to the meaning of its component words. The 'priority thesis' is encapsulated in Gottlob Frege, *Begriffsschrift*, and Jean van Heijenoort, *From Frege to Gödel: a source book in mathematical logic, 1879-1931* (Cambridge: Harvard University Press, 1977), in the theoretical priority of judgements over their constitutive elements.

keeping in sight the fact that concepts and objects occupy different position in his ontological hierarchy.⁹

Merging all three principles together, it could be said that we are interested only in ‘judgeable contents,’ they constitute our ‘meaningful units,’ and they could be further analyzed in terms of an object that falls under a concept. Thus we may get a coherent picture of the whole *Begriffsschrift*. In deploying these principles in *Grundlagen*, Frege’s strategy was to rely on the *Begriffsschrift* in a way in which it is possible to obtain a conceptual framework for analyzing the concept of number in a very logical manner, and so to fulfill the task of reducing arithmetic to logic.

1.3. The two Roles in *Grundlagen*

Let us see now what the role does CP play in *Grundlagen*. Besides its occurrence in the introduction, CP may be found in *Grundlagen* in another three places:

(§60) That we can form no idea of its content is therefore no reason for denying all meaning to a word, or for excluding it from our vocabulary. We are indeed only imposed on by the opposite view because we will, when asking for the meaning of a word, consider it in isolation, which leads us to accept an idea as the meaning. Accordingly, any word for which we can find no corresponding mental picture appears to have no content. But we ought always to keep before our eyes a complete proposition. *Only in a proposition have the words really a meaning.* It may be that the mental pictures float before us all the while, but these need not correspond to the logical elements in the judgement. It is enough if the proposition taken as a whole has a sense; it is this that confers on its parts also their content.

(§62) How, then, are numbers to be given to us, if we cannot have any ideas or intuition of them? Since it is *only in the context of a proposition that words have any meaning*, our problem becomes this: To define the sense of a proposition in which a number word occurs.

(§106) We next laid down the fundamental principle that *we must never try to define the meaning of a word in isolation, but only as it is used in the context of a proposition*; only by adhering to this can we, as I believe, avoid a physical view of number without slipping into a psychological view of it.

⁹ We may regard this point as an anticipation of the idea that ‘concepts are functions’. Another later idea will be that “everything is either a function or an object.” Since all his later elaborations are *in nuce* here, he is entitled to introduce this very Kantian dichotomy between concepts and objects.

CP has two main roles in *Grundlagen*: to reject any psychological content from logic and mathematics (§60, §106), and to introduce 'contextual definitions', required to define numbers as (abstract) objects (§62). The first role is *methodological* and stands in connection with the other two fundamental principles, whereas the other role is rather *technical*, and employs the principle as a axiom from which the theorem of contextual definition is deduced.

But if the second role is uncontroversial here,¹⁰ maybe more should be said about the connection between CP and the idea of a 'pure thought'. How can we in fact block the psychological infiltration into our logic/mathematics? Frege's response in *Begriffsschrift* was that: "to prevent anything intuitive (*Anschauliches*) from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free of gaps."¹¹ *Free of gaps* means here that once we start with pure judgeable contents, the logical system is preserving these contents, producing thus only pure thoughts. But the second step will be to secure the fact that we will be constrained to start *only* with pure contents. This is exactly the general role of CP in *Grundlagen*; since *words have meanings only in the context of a sentence*, we are throwing out the possibility of attaching independent meanings to words. Here, Frege is attacking directly the 'atomistic view of meaning', stemming mostly from the British empiricism, where words get meaning through sensorial perceptions and thus we attach to each word a mental image; our knowledge about the world is built from such images. But these images may be subjective, and thus the meanings may be subjective as well. Yet, meaning is objective for Frege, and so we need to 'purify' our mathematical thought, view which brings into discussion the role of intuition and representation in mathematics and logic. Frege is reluctant to accept the Kantian view that arithmetical truths are synthetic a priori, endorsing the analyticity of mathematics and expelling the intuition out of the mathematical realm.

Employing CP in *Grundlagen*, Frege is obtaining a secured system, where the content of the proposition is kept purely logic and this 'purity of thought' is preserved along all logical inferences. The purity of logical thought ensures us that

¹⁰ "When *Grundlagen* is read in its natural sense, without the importation of views stated only in Frege's subsequent writings, it is plain that he regarded his principle that words have meaning only in the context of sentences as justifying *contextual definitions*, and took this to be one of its most important consequences" (Michael Dummett, *Truth and Other Enigmas* (Cambridge: Harvard University Press, 1978), 95). For a detailed and interesting analysis of the role of 'contextual definitions' in *Grundlagen*, see William Demopoulos, "The Philosophical Basis of Our Knowledge of Number," *Nous* 32, 4 (1998): 481-503.

¹¹ Frege, *Begriffsschrift*, 5.

meanings are not subjective ideas, but objective contents that can be communicated and have a precise truth value.

It should be added here that CP plays even a greater role than those mentioned above, namely it marks the 'linguistic turn' in the contemporary philosophy. The language it is not further seen as a simply tool to communicate and express our thoughts, but *the* tool for approaching the world and so the analysis of language is required and prior to any other analyses. Michael Dummett claims enthusiastically that:

§62 is arguably the most pregnant philosophical paragraph ever written. (...) it is the very first example of what has become known as the 'linguistic turn' in philosophy. Frege's *Grundlagen* may justly be called the first work of analytical philosophy. (...) *There* is the linguistic turn. The context principle is started as an explicitly linguistic one, a principle concerning the meanings of words and their occurrence in sentences; and so an epistemological problem, with ontological overtones, is by its means converted into one about the meanings of sentences.¹²

2. The Interpretation of CP

2.1. The 'methodological' and 'epistemological' interpretations

How can we now interpret the principle? I think that it could be interpreted in three general ways: as a *methodological* principle, an *epistemological* principle and a *semantic* principle.

CP as a methodological principle reads as "*in order to keep pure our system*, then do not ask for the meaning of the words in isolation, but only in the context of a proposition". The *methodological* interpretation is accurate because of the methodological role of the principle in the *Grundlagen*. As we have already seen, CP is securing our logical content from any possible psychological interference. Again, the *reading key* is to keep in mind the whole project of Begriffsschrift, which was to gain a conceptual notation that will make logic the 'real science of truth'. The principle could be thus seen as operating at the methodological level, because of its capacity of providing us a way of approaching the issues. It says that from now on we have to change our habit of constructing logical proposition from the mere conjunction of subject of predicate with a new conceptual practice,

¹² Michael Dummett, *Frege. Philosophy of Mathematics* (Cambridge: Harvard University Press, 1991), 111; see also Michael Dummett, *Origins of Analytical Philosophy* (London: Duckworth and Cambridge: Harvard University Press, 1993), 5.

namely to begin the conceptual analysis with propositions.¹³ Only the propositions have real 'judgeable content,' and only after acquiring such content we can further analyze the judgement into its smaller components.

So, for Frege, we cannot speak about a *composition* of the content of the judgement from smaller contents of its component words, but rather only about the *decomposition* of the judgement into smaller parts. A judgement is a self-existent whole, which is not built of concepts, but rather the concepts are obtained by analyzing the content of the judgement. An illustrative passage in this sense can be found in a letter from 1882 to Anton Marty:

A concept is unsaturated in that it requires something to fall under it; hence it cannot exist on his own. That an individual falls under it is a judgeable content, and here the concept appears as a predicative and is always predicative. In this case, where the subject is an individual, the relation, the relation of subject to predicate is not a third thing added to the two, but it belongs to the content of the predicate, which is what makes the predicate unsatisfied. Now *I do not believe that concept formation can precede judgement because this would presuppose the independent existence of the concepts, but I think of a concept as having arisen by decomposition from a judgeable content.* I do not believe that for any judgeable content there is only one way in which it can be decomposed, or that one of these possible ways can always claim objective pre-eminence.¹⁴

On the other hand, such considerations entitle interpreters like Hans Sluga,¹⁵ Leila Haaparanta¹⁶ and Marco Ruffino¹⁷ to emphasize the reading of CP mainly as an *epistemological* thesis. The CP reads in this case as follows: "never ask about the meaning of a word in isolation, but in the context of a sentence *as expressing a judgement*, just because of the priority of judgements over their components." The context would be thus interpreted in connection with the

¹³ This idea is seen by Dummett as one of his most important and fertile ideas: "the apprehension of the central role of sentences for the theory of meaning, was one of Frege's deepest and most fruitful insights" (Dummett, *Frege. Philosophy*, 629).

¹⁴ Gottlob Frege, *Philosophical and Mathematical Correspondence*, eds. Gottfried Gabriel, Hans Hermes, Friedrich Kambartel, Christian Thiel, and Albert Veraart, trans. Hans Kaal (Oxford: Blackwell, 1980), 101; my italics in the original text.

¹⁵ Hans Sluga, *Gottlob Frege* (London: Routledge and Kegan Paul, 1980).

¹⁶ Leila Haaparanta, "Frege's Context Principle," in *The Philosophy of Gottlob Frege*, vol. 3: *Meaning and Ontology in Frege's Philosophy*, ed. Hans Sluga (New York and London: Garland Publishing, 1993), 265-279.

¹⁷ Marco Antonio Ruffino, "Context Principle, Fruitfulness of Logic and the Cognitive Value of Arithmetic in Frege," *History and Philosophy of Logic* 12, 2 (1991): 185-194.

Kantian thesis that “a judgement is prior to its constitutive elements.” That thesis is called in the literature ‘the priority thesis’, and expresses Kant’s idea that in the order of knowledge judgements are prior and only from judgements we can extract the subject-predicate relation. The epistemic unit of our knowledge of the world would be thus the judgement.

The doctrine of the priority of judgements over concepts can be understood only if it is seen as deriving from deep features of Frege’s thought. It expresses one of the Kantian elements in his thinking. Together with the Leibnizian idea of a perfect language and that of the reduction of arithmetic to logic these elements constitute the guiding principles for the construction of the *Begriffsschrift*.¹⁸

But why should we consider the judgement as the fundamental epistemic unit? The answer lies in the connection of epistemic problems (and, as we will see very shortly, semantic problems as well...) to the theory of truth. Concepts encapsulate meaning, they are meaningful, but they are not true or false. Only when connected with objects, we can speak about true facts. But, as stated above, in this case, in a purely Fregean terminology, “the objects fall under the concepts” and the recognition of that fact constitute a judgement. So, only with regard to judgements we can talk about truth and only they can be seen as the adequate truth-bearers.

Whenever we read CP in connection with the other two principles in order to reject psychologism, then we are committed to a *methodological* interpretation, whereas when we read it as restating the Kantian ‘priority thesis,’ then we are committed to a *epistemological* interpretation. They should not be seen as contradictory interpretations, but rather as complementary theses that try to capture Frege’s intentions for using CP in a very fundamental way. CP, if seen in a broader Kantian epistemological framework and along the project of *Begriffsschrift*, admits of both a methodological and an epistemological interpretation. But what if one interprets it through later writings, where Frege distinguished between sense and reference.

2.2. The ‘semantic’ interpretation

When I wrote my *Grundlagen der Arithmetik*, I had not yet made *the distinction between sense and reference*; and so, under the expression ‘a possible content of judgement,’ I was combining what I now designate by the distinctive words ‘thought’ and ‘truth-value.’ Consequently, I no longer entirely approve of the

¹⁸ Sluga, *Gottlob Frege*, 95.

explanation I then gave, as regards its wording; my view is, however, *still essentially the same*.¹⁹

So, since what was the meaning ('judgeable content') of propositions in *Begriffsschrift & Grundlagen* is now divided into sense (Sinn) and reference (Bedeutung), we may correctly wonder now whether CP is a thesis concerning only sense, or maybe only reference, or perhaps both. This line of interpretation is followed by interpreters like Michael Dummett and Michael Resnik, and I will call it the 'semantic interpretation' of CP.

Firstly, it should be made clear the point that to interpret CP as a semantic thesis does not mean at all to affirm Frege's support for some kind of 'semantic holism,' as some recent interpreters²⁰ have suggested. In this case it is not the meaning of a proposition which is 'responsible' for the meaning of its components, but a whole system of such propositions; we may thus have (that in this 'semantic holism,' what gives meaning to words and/or propositions is) either the language as a whole (Wittgenstein) or a certain theory and/or a system of such theories (Quine). But surely this was not Frege's intention.²¹

Secondly, CP implies neither that words have no meaning at all in isolation, nor that the meaning varies necessarily from sentence to sentence. The latter point means that the principle does not preclude a word to have only one meaning, whereas the former point suggests that here, in *Grundlagen*, Frege is concerned primarily with concepts and concept-words, and therefore he is not dealing with proper names, which are complete and saturated expressions, and

¹⁹ Gottlob Frege, *Posthumous Writings*, trans. Peter Long and Roger White (Oxford: Blackwell, 1979), 47; my italics.

²⁰ "If Contextuality is taken - as it has been by many - to indicate some sort of *semantic* principle, some sort of *semantic holism* whereby the meaning of individual words is constituted by, or is ontologically dependent upon, the meaning of sentences in which they occur, then there is no evidence whatsoever that Frege held the view at any time in his career, from the earliest to the latest publication and in all the unpublished works. Baker and Hacker, Davidson, Dummett, and the others who think Frege not only was a 'meaning holist' but that this is his 'most important contribution' are just wrong" (Francis Jeffrey Pelletier, "Did Frege Believe Frege's Principle?," *Journal of Logic, Language, and Information* 10 (2001): 110).

²¹ A similar position is expressed by Bar-Elli: "Does the context principle imply a kind of *holism* in the theory of meaning? Does it imply a version of *ontological relativity*, which threatens the Fregean conception of the objectivity of meanings? (...) I believe that a negative answer should be given to the first two questions" (Gilead Bar-Elli, "Frege's Context Principle," *Philosophia* 25 (1997), 100).

which seem to have senses independent from any linguistic context.²² But we should also not confuse CP with the considerations regarding the incompleteness of predicates and the completeness of proper names, which are so by their intrinsic nature and not by any extrinsic feature of a given context.²³ In the light of the sense-reference distinction, CP can be understood as two different principles:

(CPS) Only in the context of a proposition words have *senses*;

(CPR) Only in the context of a proposition words have *references*.

I will not enter in any dispute concerning which thesis is more correct, if any. I will simply say that since senses determine reference, the sense being the *mode of presentation* of the reference, it seems that whenever CPS is accepted then CPR should be accepted as well. On the other hand, if 'more correct' means here 'closer' to *Grundlagen's* claims and intentions, since there Frege is distinguishing between an objective content (judgeable content) and a subjective content (idea or mental image), it seems very natural that he had in mind the content/meaning as 'sense' and not as 'reference'; it is clear that the distinction between objects and their mental representations does not create any trouble in the sense of the problems discussed in *Grundlagen*. Thus, I will restrict myself to discuss only CPS.²⁴

²² In *Sense and Reference* we can find that: "the reference of a proper name is the object itself which we designate by its means; the idea, which we have in that case, is wholly subjective; in between lie the sense, which is indeed no longer subjective like the idea, but is yet not the object itself" (Frege, *Posthumous Writings*, 60).

²³ A clear formulation of this point can be found in Bar-Elli: "The context principle must be distinguished from the thesis that the senses of predicates and of functional expressions are *incomplete*. The latter is a much more specific thesis. This becomes manifest once we realize that if they were the same claim then Frege should have said that the sense of a name is incomplete, as that of a predicate is. The incompleteness thesis, however, is specifically about predicates, incompleteness being a feature that distinguishes them from names" (Bar-Elli, "Frege's Context," 106).

²⁴ Since the reference of a sentence is its truth-value, CPR requires a further interpretation, because to say that only in the context of a true proposition a word have reference seems somehow to reverse the natural way of dealing with truth, namely that a proposition is true exactly in the case when its constituents have references (counterparts in reality). This applicability of CPR constitutes the core of Peter Milne, "Frege's Context Principle," *Mind* XCV, 380 (1986): 491-495 analysis, and is also mentioned by Pelletier: "Frege of course does not think the *Bedeutung* of a term is a part of the *Bedeutung* of more complex expressions in which

CPS, understood as a thesis governing sense, reads as follows: *we can ask about the senses of words only in the larger context of the sense of a sentence*. But does this make sense? I think it does, in the sense that the meaning of the proposition ('judgeable content' of beginning writings and 'thought' in later works) constitutes the basic semantic unit. Why? Why sentences and not words? Exactly like in the case of epistemic reading, the complete sentence is regarded as the fundamental unit because it is the basic carrier of truth. We cannot ask about sense in isolation, outside the context of a proposition. For example we may encounter a name in isolation (like passing by a city name on the highway...), but if we are going to ask about its meaning, then we are putting it in a context, in the context of that particular thought. Thus the moral of CPS would be that whenever we are asking about the sense of a word, we are looking for it already in the context of the sense of a proposition.

CPS would become also the expression of a very fundamental insight about natural languages, namely the fact that meaning, exactly like truth, is *context dependent*. This context dependence is an intrinsic feature of its very nature. But interpreting CP as a fundamental claim about the nature of language, one seems to come in conflict with another fundamental insight, namely that natural languages have a compositional structure; we can understand the meaning of new sentences only after understanding the meanings of their component words:

It is marvelous what language achieves. By means of a few sounds and combinations of sounds it is able to express a vast number of thoughts, including ones which have never before been grasped or expressed by a human being. What makes these achievements possible? The fact that the *thoughts are constructed out of building-blocks*. And these building-blocks correspond to groups of sounds out of which the sentence which expresses the thought is built, so that the construction of the sentence out of its parts corresponds to the construction of the thought out of its parts.²⁵

This linguistic capacity of humans to understand new thoughts seems to force Frege to accept, contrary to CP, that in order to understand/grasp a new proposition we must first be able to understand the meanings of its component words. But does it mean that senses are compositional? And if so, how can we solve the conflict with CP?

it occurs. It would be absurd to think that, because "Etna is taller than Vesuvius" is true, the mountains Etna and Vesuvius are parts of The True" (Pelletier, "Did Frege Believe," 104).

²⁵ Frege, *Posthumous Writings*, 225; my italics.

This tension between the two claims is very important, because *prima facie* it seems that we have to renounce at one of the two theses. But are they in conflict? Some commentators²⁶ say yes, and in virtue of this incompatibility of the two, they are rushing to claim that Frege totally renounced at CP after writing *Grundlagen*.

Frege seems to have never endorsed explicitly CP after *Grundlagen*, but he also never acknowledged explicitly compositionality as a fundamental principle. The former does not mean either that he explicitly rejected it; on the other hand, from the latter point we cannot deduce that compositionality is not important to Frege's conception of meaning.

However, to agree that Frege changed his conception in a very fundamental way means to deny his amazing 'unity of thought'. But, since I advocate Frege's coherence, I must accommodate both features in a consistent theory of meaning, and thus to articulate an interpretation in which both contextuality and compositionality peacefully coexist. This interpretation is supported by Michael Dummett,²⁷ Gilead Bar-Elli,²⁸ G. P. Baker and P. M. S. Hacker,²⁹ or Leila Haaparanta.³⁰ Bar-Elli, for example, holds that in speaking about senses we have to distinguish between two interpretations of CP:

Let us call the first interpretation – according to which the principle tell us how to identify or determine the meaning of a term – 'the *identifying interpretation*,' the other – according to which the principle tells us in what the very idea of the meaning of a term consists – we shall call the '*essential interpretation*.' (...) The distinction between the essential and the identifying interpretations seems to me important for understanding the significance of Frege's principle, and it will

²⁶ Michael David Resnik, "The Context Principle in Frege's Philosophy," and "Frege's Context Principle Revisited," in *The Philosophy of Gottlob Frege*, vol. 3: *Meaning and Ontology in Frege's Philosophy*, ed. Hans Sluga (New York and London: Garland Publishing, 1993), 60-69, 123-137, Pelletier, "Did Frege Believe,"; on the other hand Theo M.V. Janssen, in "Frege, Contextuality and Compositionality," *Journal of Logic, Language, and Information* 10 (2001): 115-136, claims that, due the central and continuous role of CP, Frege never really endorsed something like a compositionality principle.

²⁷ Michael Dummett, "The Context Principle: Centre of Frege's Philosophy," in *Logik und Mathematik. Frege-Kolloquium Jena 1993*, eds. Ingolf Max and Werner Stelznezer (Berlin: Walter de Gruyter, 1995), 3-19.

²⁸ Bar-Elli, "Frege's Context," 99-219.

²⁹ G. P. Baker and P. M. S. Hacker, *Frege. Logical Excavations* (New York: Oxford University Press, 1984).

³⁰ Haaparanta, "Frege's Context," 265-279.

prove essential for the way I shall suggest for reconciling the apparent clash between the principle and the compositionality thesis."³¹

There is an important point of divergence between my view and Bar-Elli's. I do not consider them as two different interpretations of CP, but rather as two different ways of dealing with senses. Perhaps the distinction would be better explained in connection with the problem of truth. With regard to truth, there are two different things: the nature and the criterion(s) of truth. They response to two distinct questions: *what is truth?* and *how can we determine it?* It is one thing to define the truth and another to specify the criterions of being true.³² For instance the definition of truth can be the *correspondence* with facts, whereas the criterion would be the *coherence* among propositions. The same distinction seems to work for sense. The answer to the question *what is sense?* may be obtain by employing CP, whereas the appeal to compositionality thesis could serve us to answer to the question *how do we determine sense?*³³

But his point may be undermined by saying that, since the distinction definition-criterion of truth is not clear and without problems, the analogy may cause more problems than clarifications. Thus the reconciliation is in danger and we need a firm terrain to build up a common accommodation of the two claims. An important insight for this reconciliation lies in Dummett's slogan that "in the order of *explanation* the sense of the sentence is primary, but in the order of *recognition* the sense of a word is primary."³⁴ This thought captures precisely the nature of the two apparently contradictory points. When we ask for the nature of the sense, for a theoretical *explanation* of what meaning is, then the role of CP is exactly to make clear the point that sentences are prior to words, and they should be considered as complete sense carriers. On the other hand, when we try to see

³¹ Bar-Elli, "Frege's Context," 103

³² The distinction is explicitly stated in Russell: "coherence cannot be accepted as giving the *meaning* of the truth, though it is often a most important *test* of truth after a certain amount of truth has become known" (Bertrand Russell, *The Problems of Philosophy* (New York and Oxford: Oxford University Press, 1997), 123).

³³ In the light of the previous discussions, the connection with truth here is not *ad hoc*, but it follows the intimate connection between truth and sense. In Frege's semantics truth and sense are deeply interconnected. As Dummett points out: "to grasp the sense of a sentence is, in general, to know the conditions under which that sentence is true and the conditions under which is false" (Dummett, *Frege. Philosophy*, 5).

³⁴ Dummett, *Frege. Philosophy*, 4; for further elaborations of this point see also Michael Dummett, *The Interpretation of Frege's Philosophy* (Cambridge: Harvard University Press, 1981), 547.

the criterion of being meaningful, of how we are actually grasping senses, then we are looking for something else, namely for a *recognition* of how the things works in this case.

This semantic picture³⁵ resembles very much with the Leibnizian metaphysical view concerning the part-whole relation, where the parts are prior to the whole in the case of actual discrete objects, whereas in the case of continuous ideal objects the whole is prior to its parts. Of course that the smallest meaningful carriers of sense of natural languages are words, yet we learn words and use them in order to produce sentences, like we produce bricks not for themselves, but in order to put them together and build houses. It is like in molecular chemistry: we acknowledge the existence of submolecular levels like atoms, electrons, quarks and so forth, yet the theoretical level of analysis is set at the level of molecules. They are relevant for our investigation, even though they are made up of various combinations of atoms. The comparison is further relevant for in nature as in natural languages, we very rarely may find solitary atoms; most of them come up combined in molecules. Molecules made up the surrounded universe, even though they are in fact composed of atoms. So, both contextuality and compositionality could peacefully and fruitfully coexist under the same Fregean roof.³⁶

What needs perhaps here to be added is the fact that all the three interpretations of the principle should be seen as complementary to each other,

³⁵ This is also similar with Socrates' talk about 'wholes' in *Parmenides*; we can regard either the whole as divisible into parts or the parts as forming up the whole.

³⁶ All this Fregean problematic issues seems to have its echoes in *Tractatus*, where both contextuality and compositionality are to be found:

- Contextuality:

3.3. Only propositions have sense; only in the sense of a proposition does a name have meaning.

3.314. An expression has meaning only in a proposition. All variables can be constructed as propositional variables.

- Compositionality:

3.318. Like Frege and Russell I construe a proposition as a function of the expression contain in it.

4.026. The meanings of simple signs (words) must be explained to us if we are to understand them.

Wittgenstein's later conception of the meaning of a word as its use in the language (games), could be regarded as a 'mere' extension of Fregean CP. For a detailed and interesting analysis of this point, see Erich H. Reck, "Frege's Influence on Wittgenstein: Reversing Metaphysics via the Context Principle," in *Early Analytic Philosophy. Essays in Honor of Leonard Linsky*, ed. William W. Tait (Chicago: Open Court Publishing Company, 1997), 123-185.

rather than mutual exclusionary. I do not think of this classification as bringing into light distinct Fregean views and thus any overlapping zone among the three points is excluded *ab initio*. I rather see these interpretations as a natural succession of views, starting with the broadest interpretation and ending with the narrowest. A methodology gives one a way of approaching things, epistemology restricts this way only to the realm of knowledge, and semantics preserves from knowledge only the parts relevant to meaning. The link between the last two points can be even more explicitly exhibited by the slogan that “a theory of meaning is a theory of understanding,” and since to understand something means to know it, the connection would be obvious in this case.

OBSERVATION AND INDUCTION¹

Theodore J. EVERETT

ABSTRACT: This article offers a simple technical resolution to the problem of induction, which is to say that general facts are not always inferred from observations of particular facts, but are themselves sometimes defeasibly observed. The article suggests a holistic account of observation that allows for general statements in empirical theories to be interpreted as observation reports, in place of the common but arguably obsolete idea that observations are exclusively particular. Predictions and other particular statements about unobservable facts can then appear as deductive consequences of such general observation statements, rather than inductive consequences of other particular statements. This semantic shift resolves the problem by eliminating induction as a basic form of inference, and folding the justification of general beliefs into the more basic problem of perception.

KEYWORDS: observation, induction, problem of induction, Karl Popper, hypothetico-deductive, foundationalism

In this article, I offer a simple technical resolution to the problem of induction, which is to say that general facts are not always inferred from observations of particular facts, but are themselves sometimes defeasibly observed. I suggest a holistic account of observation that allows for general statements in empirical theories to be interpreted as observation reports, in place of the common but arguably obsolete idea that observations are exclusively particular. Predictions and other particular statements about unobservable facts can then appear as deductive consequences of such general observation statements, rather than inductive consequences of other particular statements. This semantic shift resolves the problem by eliminating induction as a basic form of inference, and folding the justification of general beliefs into the more basic problem of perception.

In the first section of the paper, I analyze the problem of induction in terms of five jointly inconsistent propositions, of which the weakest is the statement that

¹ I would like to thank Earl Conee, Richard Fumerton, Alan Sidelle, Elliott Sober, audiences at SUNY-Geneseo and the Creighton Club, and several anonymous reviewers for helpful comments on earlier versions of this article.

all observations are particular rather than general. In the second section, I complain about the standard particularistic theory of observations, which depends on a cluster of assumptions that are commonly taken for granted, but that deserve little support in the light of recent progress in philosophy. In the third section, I give a brief sketch of a possible holistic account of observations, and show how it might work as a positive solution to the problem. I suggest that a main weakness in the classical hypothetico-deductive model of scientific reasoning can be removed if at least some hypotheses can be seen as defeasible observations of general facts.

Let me be clear about what I think I can establish. My primary concern is to point out that there is a possible new approach to the problem of induction in terms of general observations—an approach that ought to be considered, but is somehow missing from the standard treatments of the issue. My secondary concern is to argue that there really are such general observations. I do not want the value of this essay to depend entirely on that idea's being independently more plausible than other theories about observation. I am not certain that it is. But if it has any plausibility at all, and if it really gives us a way to resolve the problem of induction, then it will be worth some future effort to work the idea out in detail.

I. The problem of induction

An inductive inference is often defined as one in which the conclusion does not follow necessarily from the premises—so it is not deductively valid—but in which the premises seem to render the conclusion more likely.² This is sometimes seen as a matter of the conclusion's somehow adding to the content of the premises. As Brian Skyrms puts it, “If an argument is inductively strong, its conclusion makes factual claims that go beyond the factual information given in the premises.”³ Wesley Salmon calls anything like this an ‘ampliative’ inference.⁴ (E1) and (E2) below are simple examples of these ampliative inferences.

² I will concentrate on one standard type of definition of induction, convenient for my purposes. I believe that what I say can be extended to apply to other common formulations, but will not attempt to do so here. James Cargile provides a discussion of various definitions in “The Problem of Induction,” *Philosophy* 73 (1988): 247-275.

³ Brian Skyrms, *Choice and Chance* (Belmont: Wadsworth, 1986), 8.

⁴ Wesley C. Salmon, *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1967).

- (E1) This raven is black.
That raven is black.
 All ravens are black.
- (E2) All ravens observed so far are black.
 All ravens are black.

A third common form of inductive argument moves from what is known or observed to particular unknown cases, for example:

- (E3) All ravens observed so far are black.
 The next raven observed will be black.

This third form may be seen as deductive extension of form (E2), since if we take our observations to imply some general fact, then we can also take them to imply whatever is entailed by that fact. It might also be seen by some as having independent standing as a form of inductive argument. In any case, I will concentrate on forms (E1) and (E2) in what follows. These examples best fit Karl Popper's largely syntactic understanding of induction:

It is usual to call an inference 'inductive' if it passes from *singular statements* (sometimes also called 'particular' statements), such as accounts of the results of observations or experiments, to *universal statements*, such as hypotheses or theories.⁵

The conclusions of (E1) and (E2) do not follow necessarily from their premises, evidently because the conclusions say more than the premises, in that they talk about all ravens, not just those mentioned in the premises. The problem of induction is, then, often understood to be the problem of justifying non-deductive inferences like these.⁶ As Hume was the first to point out, since such

⁵ Karl Popper, *The Logic of Scientific Discovery* (New York: Basic Books, 1959), 27.

⁶ This is controversial. There are many who would like to believe in some kind of ampliative inference, but who also think that the little forms listed are worthless in themselves. We know that the sun will rise tomorrow, not simply because we have a series of past risings of the sun; there must be something else involved, that distinguishes the law-like regularities from the merely accidental ones. A recent strategy attempts to replace enumerative induction with abduction or "inference to the best explanation" (see Hilary Putnam, "The Meaning of Meaning," in his *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975), 215-271). I am inclined to agree with Richard Fumerton, in "Induction and Reasoning to the Best Explanation," *Philosophy of Science* 47 (1980): 589-600, that this form of reasoning is

inferences cannot be justified deductively, and cannot be justified inductively either (on pain of circularity), it appears that they cannot be justified at all.

Why should we care about the problem of induction? The answer is that we seem so heavily to depend on such inferences, in science and in ordinary life. That is, we accept as justified many beliefs that can be viewed as the conclusions of inductive inferences, and we further believe that such beliefs originate in inductive inferences. If no such inferences are rationally justified, it looks like we ought to give up much of what we now believe.

Why do we think that what might be called “inductive conclusions,” such as that all ravens are black, require inductive arguments? Perhaps because we are empiricists, in at least the broad sense that we believe (or would like to believe) that there are two and only two basic ingredients in human knowledge: observation and proper reasoning (where by proper I mean valid, or else rationally justified in some other way). It may be that we can figure out some things, such as truths of mathematics, *a priori*, through valid reasoning alone. But our knowledge of such things as ravens is not like that; it must be based on observation as well. Unfortunately for general beliefs, it seems that all we can observe at any one time is this or that raven (or, at most, some small number of ravens) and their properties. The general statement that all ravens are black is not deducible from any available set of reports of observations about particular ravens, though those are all that we have to go on. This is why we have a problem, and why it looks as if we need to find some way of justifying ampliative arguments. But I want to reconsider the implicit claim that the general facts in question are always unobservational. I want to suggest that we come to believe them in essentially the same way that we believe particular facts, and with the same kind of justification.

The distinction that I will employ between general and particular statements, facts, or observations is not identical to Popper's, and needs a more definite characterization. There are three types of statements that we usually find listed as the premises in inductive arguments. Some are singular claims of the form “this A is B” or “the C A is B,” such as “this raven is black” or “the twelfth observed raven is black.” Others are existential claims of the form “Some A's are B,” “A least two A's are B,” and the like. And still others are universal statements of the form “all C A's are B,” such as “all of the ravens in such-and-such a sample

effectively reducible to induction. If I am wrong, and abduction must be seen as a distinct form of ampliative inference, it nevertheless stands in the same need of justification as induction. What I say in this paper may be applied as well to the resulting “problem of abduction” as to the traditional problem of induction.

are black,” or “all observed ravens are black.” It appears that none of the statements usually used as inductive premises have the simple form “all A's are B.”⁷ This seems a contingent, language-dependent feature of ordinary observation reports. We could always introduce a term like ‘obsraves’ to denote the class of ravens that have been observed, and then produce the simple universal statement “all obsraves are black.” We could also artificially produce a statement like “all ravens are unobserved-or-black.” But given the way that we normally speak, it appears that the usual inductive premises *about A's* are effectively particular, in the sense that none of them affirms anything straightforwardly about the entire class of A's, but only about some members, or about a certain subclass.

I will call any contingent statement that is effectively particular in normal language in the way that I have described a *p-statement*. I will call any statement that takes the form of a simple universal affirmative sentence a *u-statement*. In what follows, I will call the facts (if they exist) to which p-statements and u-statements correspond p-facts and u-facts. I will call the objects (if any) to which the subject terms of those statements refer p-objects and u-objects. And I will call observations (if they occur) of p-facts and u-facts p-observations and u-observations. My point is just to focus on the kinds of statements that are involved in alleged inductive inferences, as distinct from the epistemic roles that these statements are supposed to play.

Now I can summarize my understanding of the problem of induction as a set of five jointly inconsistent statements:

(S1) Our knowledge (or justified belief) has the form of a set of observation reports and their consequences closed under proper inference.⁸

(S2) All observation-reports are p-statements.

(S3) All proper inferences are deductive.

(S4) It is impossible to deduce a u-statement from any set of p-statements.

(S5) We have knowledge (or justified belief) of the truth of some u-statements.

⁷ An exception would be “All of my fingers are unbroken,” or something of the sort, where one knows that the entire relevant class is present to the observation.

⁸ The class of analytic propositions should be included as well, if these are considered to be substantive objects of knowledge.

Any reasonable approach to the problem of induction must falsify at least one of these five statements. To reject (S5) would be to embrace skepticism with respect to the whole class of universal statements. This is a possible view, of course, but not what we should call a solution to the problem.

Statement (S4) is hard to deny. I cannot prove that it is true, for the obvious reason that the classes of u- and p-statements are only partly defined. But it is demonstrably true for the standard cases that I have in mind—for example, no proposition of the form “all A’s are B” can be deduced from any set of propositions of the forms “this A is B” and “all C A’s are B.”⁹

In most standard presentations of the problem, such as Salmon’s, it is simply presupposed that something like statement (S3) must be rejected if the problem is to admit of a solution. There have been many attempts to prove that one or another non-deductive inference pattern is proper. None of these efforts has gained very wide acceptance. Popper and other deductivists affirm (S3) and treat inductive inference as an illusion, arguing that science works essentially through the falsification of some tested hypotheses. But this leaves the positive justification of surviving hypotheses problematic.

(S1) is intended as a concise statement of the central claim of empiricism. While it is surely subject to objections and qualifications, few traditional philosophers of science would deny it wholesale or in spirit. This does not entail that (S1) is true, of course. My point is rather that induction is primarily a problem for broad-sense empiricists in the first place.

There is room in this analysis for another approach to the problem: Deny statement (S2) above. Assert in its place that ordinary u-statements like “All ravens are black” can sometimes be accepted as reports of observations, or as deductive consequences of more general u-statements that are reports of observations. This approach could give us a quick, snappy solution to the problem of induction, if it did not seem so obviously to be false. I want to say that it is actually true, despite appearances—or, at least, that it can be treated as true for purposes of philosophical analysis. In what follows, then, I will do what I can to make the idea of non-particular observations less implausible. To that end, I will try to undermine the common assumptions that support (S2), and to replace them with a quick sketch of an alternative theory of observation. The result will sympathize with Popper’s rejection of induction as a fundamental form of reasoning, but offer the idea of general observations as a positive means of justifying “inductive conclusions.”

⁹ I am ignoring the possibility that C is a vacuous property like ‘self-identical.’

II. The common theory of observation

Why does it seem so obvious that all observations are particular? The claim that only p-facts can be observed is not essential to broad-sense empiricism. It stems, rather, from a certain theory about observation. This theory has its roots in common sense, to be sure, and has appeared in philosophical writings since Aristotle's *Posterior Analytics*. But its largely unchallenged status in epistemology may stem more from convenience and simplicity than from any claim to universal truth. It is, in fact, a theory of observation that most present-day philosophers will cheerfully reject when it causes problems in other contexts.

According to the common theory, the philosophically best cases of observation are quite local and brief, such as an individual person's seeing that a certain object in his presence has a certain color. These quick, individual observations find their most natural expression in the form of p-statements. All other cases will be seen as proper observations only to the extent that they approximate these paradigms. This view of observation accords well enough with pre-philosophical intuitions. It is obvious that we can't see everything at once, and we can surely see things better when they are nearby and reasonably small. But for this idea to function as a philosophical theory of observation, not just a rule of thumb, requires further metaphysical, semantic, and epistemological assumptions.

There are three most important such assumptions, and all three have been losing force within philosophy over the past several decades. The first assumption is that, since observational beliefs are epistemically foundational, they should be absolutely certain, or at least as close as possible. The second is that knowledge and justified belief ought to be seen as existing primarily or exclusively in individual minds. The third is that discrete individual objects and their properties are fundamental to the metaphysical and semantic structure of the world. All of these common assumptions were important to the positivists' original project of rationally reconstructing scientific knowledge within something like a classical first-order logical language. Absent the requirements of that project, however, the claim that only particular, immediate facts are observable can be at least reopened for discussion among broad-sense empiricists. Let me reconsider the three background assumptions of the standard theory, then, one at a time.

It used to be held that observations, or at least a certain foundational class of them, must yield absolutely certain knowledge. But few philosophers think this way anymore, and it was never very plausible to apply that criterion to ordinary reports of observations, as distinct from artificial statements about sense-data. For example, if I think I see that a particular raven is black, I can be wrong in a number of ways. It could turn out to be a big crow, not a raven. It could be navy

blue, not black. It could be black on the side facing me, but pink on the other side. I could even be dreaming or hallucinating the whole experience. If we are to speak about ordinary objects rather than immediate sense-data, we can say at best that observing (or seeming to observe) a particular fact gives us good, *prima facie* reason to believe in that fact, but nothing more. As we now say, observational beliefs are *defeasible*. With additional observations and reports from other people (in case there's something wrong with our own eyes, for example), we might get closer to certainty, though we will never get all the way. But if there is no special need for certainty, if all we require of observation is that it give us *prima facie* justification, then there is less reason to restrict the scope of observation to local facts and objects. If I can report, defeasibly, the observation that a certain Roman driver ran his motorcycle into a certain pedestrian, why can I not report defeasibly the observation that Romans in general are reckless drivers? Neither is certain on its face; both would require further investigation to pronounce as definitely true. And many American tourists do claim to observe the general fact that Romans are reckless drivers, calling it an observation in the ordinary sense of the word, just as they claim to observe this or that particular collision or near miss. It is not clear that there is any philosophically essential difference here.

Traditional empiricists have also worried about skepticism with respect to memory. If we believe in foundational observations, we can only get around the problem of memory by requiring that those observations be discrete and very brief events—too brief for memory to play an internal role in the process. Bertrand Russell's remark to the effect that sense-data last "about two seconds" is sometimes seen an amusing example of philosophical bullet-biting. But why does this straightforward statement strike us as funny? I think it is because everybody knows that observations are the sort of thing that can be individuated only arbitrarily. As we speak about them outside of philosophy, observations are often highly indeterminate in duration and scope. Two seconds may actually be an approximate lower bound of sorts: it is about the length of time it takes per sentence to make a series of oral reports at top speed, like a play-by-play announcer at a football game. But this is hardly significant for epistemology. Nor is it relevant that it takes something like a tenth of a second for a person to notice any particular change in his surroundings, since those intervals are not discrete, but plainly overlap each other in a more-or-less continuous way. And unless we wanted to maintain that perception was infallible, while memory was not, there would be no good reason to be concerned about such lower bounds in the first place.

As we usually speak, events and processes of all durations can be observed, and those observations reported. A person can say that he has seen the sun set, seen a new bridge go up, seen an army lose a war, and so on. Why should any of these things be ruled out as proper observation reports? If even the rise and fall of the Third Reich can be considered as one big event, comprising lots of particular and general facts, we should be able to describe William Shirer's lengthy book of that title as the report of one big observation: this very big thing happened, he watched it happen, and the book is his report.

A second background assumption to the standard theory of observation, hence to the problem of induction, is what is called methodological individualism, or sometimes, rather pejoratively, methodological solipsism. This has also been widely rejected in recent decades. It has one source in traditional concerns about the problem of other minds. If, as above, we are determined to base our beliefs on a foundation of certainty, and if the existence of other people's minds is impossible to establish, then we can hardly grant the observations of others equal status with our own. This results in the restricted view that each person's knowledge must be based solely on the observations that he is able to make for himself.

But again, it is not clear that we ought to impose this limitation on the range of observable facts. In ordinary life, we often take reports of others' observations (for example, those of our parents or doctors) as perfectly good grounds for our own beliefs.¹⁰ Moreover, we frequently make reports of shared observations, speaking in the first person plural. (For example, the previous sentence.) Observation reports are given by teams of researchers, by businesses and government agencies through their public relations offices, and by all sorts of other groups.¹¹ Consider also Hilary Putnam's discussions about metals and trees.¹² Most of us know many things about aluminum, he says, for example that it's cheap and shiny, without being able to distinguish the stuff from molybdenum, or any number of other metals, face-to-face. This implies that our even knowing what we are talking about, in some cases, relies on the existence of distant experts who could make the meanings of our statements more precise. In general, it is

¹⁰ I have argued elsewhere that such deference is rationally required of us in a very broad range of cases (Theodore J. Everett, "The Rationality of Science and the Rationality of Faith," *Journal of Philosophy* 98 (2001): 19-42), and that it is through such rational acceptance of the statements of others that we come to know that other minds exist (Theodore J. Everett, "Other Voices, Other Minds," *Australasian Journal of Philosophy* 78 (2000): 213-222).

¹¹ John Hardwig gives an example of a scientific paper with 99 co-authors, in "Epistemic Dependence," *Journal of Philosophy* 82 (1985): 335-349.

¹² Putnam, "The Meaning of Meaning," 225-227.

increasingly clear that much of human knowledge is distributed socially, rather than duplicated inside each of our heads. If this is right, then there should be no harm in our accepting at least some groups as capable of making at least some observations. The larger are the groups of people who can act together as observers, the bigger and more broadly scattered are the facts and objects we should take as minimally observable.

Suppose I want to say that central planning in agriculture always reduces output. I might describe this as an inductive conclusion of my own, based mainly on written sources, most of which are based on other testimony, books, reports, and scholarly analysis. But there is no reason that this general statement could not be classified as an observation that people have made collectively, rather than an inductive conclusion that I have drawn individually. Statements about well-known facts are often phrased this way in literature, to indicate points that are taken for granted by the writers and their readers. Thus, "...we have seen that it is the Holy Spirit who brings about the wonderful communion of believers in Jesus Christ,"¹³ and "...we have seen that no religion stands on the basis of things known... so must it ever be at once a source of error and contention,"¹⁴ and innumerable similar statements.

A third obsolete assumption that supports the traditional theory of observation is logical or metaphysical atomism. The broad idea is that there is one basic level of objects or properties in the world, and that everything else is analyzable in terms of these simplest items. For the early logical positivists, this was a matter of fitting the world to the structure of first-order logic and set theory. Since the collapse of the positivist project in the mid-20th century, almost nobody now thinks that classical logic is adequate to mirror the structure of the world or to analyze scientific discourse. For those who saw the world as fundamentally a set of what I am calling p-objects or p-facts, a particularistic theory of observation was only natural: if there are not really any u-facts or u-objects to begin with, if such things are only logical constructs, then there is nothing special for a u-observation to report. But most of us now hold a less restricted view of the relation between particular and general things. Some find it better, for example, to view the relation of individuals to kinds (e.g. to species in

¹³ Pope Benedict XVI, in a speech at the World Youth Day Vigil, held in Australia in 2008.

¹⁴ Francis Wright, "Morals," in *Course of Popular Lectures* (BiblioLife, 2009), 108.

biology) as more like the intrinsic relation between parts and wholes than like the formal relation between members and sets.¹⁵

There may also be other reasonable choices for the form of an observation report than particular and universal statements as they are classically understood. For example, an improved, non-atomistic semantics might be able to provide an adequate analysis of generic statements.¹⁶ Why shouldn't we say that we have observed the fact that "ravens are black," where the word "ravens" can be understood as picking out the species, in the way that the phrase "this raven" picks out the individual? "All ravens are black" might then be seen as fundamentally similar to "All of this raven is black." Each refers to a certain piece of the world, and says that the entire piece is black.

It could be objected that a *causal* theory of perception favors particularism, in that only a small number of ravens can ever figure causally in any act of observation. But it is not clear that this is true. If the part-whole idea is to be taken seriously, it may be correct to say that whenever particular ravens are involved in an event, ravens in general are also involved, just as an observation or some other event involving one room in my house necessarily involves my whole house too. Moreover, the objection presupposes an atomistic view of the entire causal situation: particular light bouncing off of particular ravens into particular eyes. But there are causal facts at macroscopic levels, too. Unless we are still trying to work within something like positivist limits, nothing prevents us from talking about light *in general* bouncing off of ravens *in general* into the eyes of people *in general*.

We often do use generic statements, rather than u-statements, to report our observations of general facts, and we do so for practical reasons. We are all concerned that our statements reflect, if not full certainty, at least a reasonably high degree of confidence in what we report. Unless we are deliberately engaged in philosophical or scientific theorizing, it is ordinarily safer simply not to report our u-observations as such—that is, not to "generalize" unduly, even if what we are observing is a universal fact. One alternative is just to report those p-observations that we are making at the same time, as we do in scientific lab reports, since these are in practice less likely to be defeated later on. The other is

¹⁵ See, for example, David Hull, *Philosophy of Biological Science* (Englewood Cliffs: Prentice-Hall, 1974), 48f, and David Sloan Wilson and Elliott Sober, "Reviving the Superorganism," *Journal of Theoretical Biology* 136 (1989): 337-356.

¹⁶ For a collection of recent efforts, see Gregory N. Carlson and Francis Jeffry Pelletier, *The Generic Book* (Chicago: University of Chicago Press, 1995).

to use the generic form of statement instead, which hedges on the possibility of some indeterminate number of exceptions (not necessarily a minority) to the universal claim. These statements are vague, obviously, but not inherently more vague than ordinary singular statements. In both cases, the subject term picks out some object in the world (say, ravens in general, or some particular raven or group of ravens), and the predicate is used to say something about it. In neither case is it strictly entailed that all, or even most, parts or instances of the subject have the property predicated of the subject as a whole. What is entailed is only that *enough* of the subject has the predicated property. The appropriate sufficiency conditions are not implicit in the statements themselves.¹⁷

Universal statements like “all ravens are black” are more precise. Such statements correspond to the world in the same way as do those that could be called *universalized* singular statements, such as “all *of* this raven is black.” The subjects are again things like an individual raven or ravens in general, but the word “all” has the function of applying the predicate to exactly all, not merely enough, relevant parts or instances of the subject. We can imagine ordinary singular and generic statements as opposite ends of a spectrum, with subject-predicate statements about mass-type objects (which are often thought of as

¹⁷ E. J. Lowe has made a partly similar, but to my mind needlessly subtle, suggestion, in “What is the ‘Problem of Induction’?”, *Philosophy* 62 (1987): 325-340. Lowe claims that the class of what I am calling inductive conclusions should not be formulated as u-statements in the first place, but rather as generic statements, which express *laws*, as he understands them, rather than universal generalities. Lowe does not quite say that these law-like facts about biological species and other kinds are themselves observable, but rather claims that observations of their “normal” instances are strong *prima facie* evidence of their truth. This is an attractive view, but hard to evaluate because the concept of a law is so elusive. For one thing, Lowe notes that in order for his laws to count as useful knowledge, we must be able to draw predictions from them in a justified way. But how, for example, can we draw “this is black” from the premise “this is a raven” and the generic formulation “ravens are black?” Not deductively, as Lowe concedes. He relies instead on the principle that most members of a kind must be normal members, so that we can make this sort of inference, in effect, probabilistically. Lowe sees the principle as analytic – it is “incoherent”, he says, to suppose it false (Lowe, “What is the ‘Problem,’” 336). But one can easily imagine cases where most of the actual instances of some type are abnormal. For example, some new plague or political development could bring it about that the majority of Canadians have no teeth, without falsifying the claim that a normal (as distinct from average) Canadian does have teeth. More recently, both Howard Sankey (“Induction and Natural Kinds,” *Principia* 1 (1997): 239-254) and Brian Ellis (“An Essentialist Perspective on the Problem of Induction,” *Principia* 2 (1998): 103-124) have approached the problem of induction along the same broad lines as Lowe, through consideration of the essential properties of natural kinds. Both stop short of claiming that the relevant facts are observable.

“scattered particulars”) in the middle. At one end of a parallel spectrum would be ordinary u-statements, and at the other end of that spectrum would be universalized singular ones. I do not know what it would take to prove that these connections are as real and as gradual as I suggest. But perhaps these features can be *observed* in the following matrix of statements:

	<u>simple (s/p)</u>	<u>universal</u>
<u>general</u>	Ravens are black.	All ravens are black.
	Apples are red.	All apples are red.
	Peas are green.	All peas are green.
	Pease is green. ¹⁸	All pease is green.
	Corn is yellow.	All corn is yellow.
	Snow is white.	All snow is white.
	The snow is white.	All (of) the snow is white.
	The sky is blue.	All (of) the sky is blue.
	The moon is silvery.	All of the moon is silvery.
<u>particular</u>	This raven is black.	All of this raven is black. This raven is all black.

The statements in each column are similar in form. The subjects get less ‘classy’ and more ‘massy,’ then less ‘massy’ and more individual as we move down the page. My claim is that these differences are not very important from an epistemological point of view, unless we are already committed to an atomistic analysis.

Atomism skews the sample for the problem of induction. It forces us to take the most particular singular statements as paradigmatic observation-reports, and to wonder how we get from them to the least particular universal statements. It is more reasonable to take all subject-predicate statements (including generics) to be equally possible reports of observation, and then to ask how they all relate to the

¹⁸ ‘Pease’ is an archaic mass noun for peas, as in “pease porridge hot, pease porridge cold, pease porridge in the pot, nine days old.”

corresponding universal statements. Residual problems about confirmation should be the same in principle for the most particular cases as for the most general.

III. An alternative theory of observation

Here is the main idea for an alternative, holistic theory of observation. Think of the world not as a set of pre-cut facts, but as a single, variegated but undivided object. Think of experience not as a series of pre-cut, sentence-like events, but as a more-or-less continuous flow that needs interpretation to be represented propositionally. Think of single experiences as non-random chunks of this whole flow of experience, unified under a broad range of possible criteria. Think of observations as articulate representations of experiences, expressed as statements. On this view, an observation could be large or small, brief or enduring, individual or social. When someone says "I see that your dog is wearing trousers," this expresses a particular observation that fits the standard subject-predicate model, made by an individual more or less momentarily. When someone says "We see that solar activity influences climate," this expresses a general observation, made not individually but socially, and very extensive in time and space. Both are legitimate sorts of observations, because the world has larger and smaller parts, and our experience has larger and smaller parts to match.

There are no *a priori* limits on what sort of empirical theory might best represent our total experience. Therefore, any amalgam of individual or collective experience could theoretically count as an observation, and any statement could count as an observation statement. Ultimately, our decisions as to what to count depend of how our total experience is best systematically articulated into a theory about the whole world. Proximately, though, we do need to rely on rules of thumb regarding what to count provisionally as observation and observability. What I am doing here, then, is debating the restrictive rules of thumb currently in use, and suggesting a more open approach as helpful to philosophical analysis, if not to practical science. I say that we have insufficient reason to insist that one syntactically-defined subset of beliefs is based on observation alone, and the rest only on inference. I think that no belief should be seen as either purely observational or purely inferential. All are functions of a total process that takes in information from the world at various levels of generality, framing hypotheses from these observations, deducing consequences, testing, taking in more observations, and gradually forming an articulate and stable model of the whole, complex system.

Even in the case of an individual observer having a very local experience over a short time, there is no essential particularity in the experience itself. There

is no difference in the initial set of sensations between those representing the blackness of ravens in general, for example, and those representing the blackness of this raven in particular. The psychological content of an ordinary observation is not very much like a sentence, after all. From the subjective or internal point of view, we begin with an experience, i.e. *some* experience, and that experience may bring some sentence or sentences to mind. We may or may not articulate that experience with such sentences, but the experience itself is something else. In reality, our observational life is much more like a *flow* of initially inarticulate sensations than it is like a series of sentences being fed in through the senses like input to a computer. Nothing prevents our expressing some of that flow of experience in general terms. We may come into a certain stream of impressions that is both ravenly, as it were, and black. We may then articulate these impressions in an appropriately vague particular form ("this raven is black") or generic form ("ravens are black"), or both. But then to universalize these simple, subject-predicate reports requires something else, a decision that sufficient evidence exists to count the object in question as consistent in all of its parts. We may need to examine more of this raven to conclude that all of it is black, or to examine more of the species raven to conclude that all of them are black. How complete these further tests must be depends on the level of certainty that we require for the resulting universalized beliefs.¹⁹

Moreover, when we think of observations taking place over longer periods of time (such as a detective's observing that a staked-out gangster always visits a certain nightclub at about one in the morning), all the less does it seem like importing a sentence through the eyes, and all the more like the selection or creation of a sentence to articulate some feature of an otherwise unseparated mass of impressions. And the more so still, when we consider that some observations might be scattered over many persons, as with a group of veterinarians and

¹⁹ It is also possible to construct or interpret empirical theories without including definite judgments as to the truth of any particular or universal statement. Instead, we can associate each statement with a probability, and let those probabilities rise and fall according to new evidence, but never reaching either 0 or 1. Bayesians consider a certain formulation of this idea, using Bayes's Theorem in the probability calculus to govern changes in subjective probabilities, definitive of empirical rationality. Wesley Salmon makes the case for this view in "Rationality and Objectivity in Science or Tom Kuhn Meets Tom Bayes," in *Philosophy of Science: The Central Issues*, eds. Martin Curd and J. A. Cover (New York: Norton, 1998), 551-593. Clark Glymour argues against it in "Why I am not a Bayesian," in his *Theory and Evidence* (Princeton: Princeton University Press, 1980), and also in *Philosophy of Science: The Central Issues*, eds. Martin Curd and J. A. Cover (New York: Norton, 1998), 594-606.

ranchers who collectively perceive an outbreak of mad cow disease in their vicinity. To put it sweepingly: there is a whole subjective world, in complex, causal contact with the whole objective world. This contact produces (or possibly constitutes) a mass of evidence. This evidence is then cut up in various ways for various purposes, with appropriate degrees of generality, from one baby seeing one red ball, to a team of scientists observing the long-term effects of a drug on tumors, to humanity as a whole discovering that cooked meat is easier to chew.

I am relying, plainly, on a certain broad faculty of choice, which is involved in our deciding how to aggregate or individuate ourselves as the subjects of our observations, how to aggregate or individuate the objects of our observations, and how to articulate the content of the resulting evidential mass. But we cannot just say whatever we want; there are important constraints that must be placed on any plausible theory of observation. It must be possible, for one thing, to distinguish good observations from bad ones. It must also be possible to distinguish what is observable in principle from what is not. And both theoretical distinctions must accord reasonably well with common intuitions.

First, then, a theory of general observations must leave room for mistakes. It must be possible to distinguish a real general observation (i.e., a correct observation of an actual general fact) from an apparent observation of a general fact that does not exist. For example, if we can observe the fact that all ravens are black (which I have been taking to be true) while directly confronting only some of those ravens, then why do we not properly observe that all swans are white (which is false), when confronting a similar number of white swans? Such mistaken general observations will have to be understood in the same way that we understand mistaken particular observations. I may see a blue car from a distance and perceive that it is blue, in which case I have observed that fact correctly, but I may also see a green car as a blue one, in which case I have made a mistake. In many cases, I may not be able to tell the difference without further research. As I said above, even a single raven in my hand may appear to me to be black, but turn out to be navy blue, or to be pink in those parts I am not directly looking at. We would still say that if it is black, then I am seeing that it is black, not inferring that it is black. This is true, even though my ability to see the raven as a whole relies on the truth of my assumption that the partial surface that I directly see is fairly representative of the entire raven.

Next, the new theory must also preserve something of the intuitive distinction between observable and unobservable objects, facts, etc. This can be done, I think, along the same lines. What is observable in the new view will be any object or fact, particular or not, in an appropriate relation to the observer,

individual or not. Presumably, this will include such general objects as the species raven (which is observed along with its instances, like every other natural kind), and such u-facts as that all ravens are black, as well as such p-facts as that this or that observed raven is black. But it will necessarily exclude those specific facts and objects which are entirely unobserved, such as the species Martian, or the fact that this or that unobserved raven is black, or that all ravens after the year 2500 are black, or that all Martians carry swords.

This may seem to generate a bit of a paradox, in that I am classifying some general facts as observable while some of their deductive consequences are not. If we have observed that *all* ravens are black, how can it be sensibly said that we have not observed that *each* raven (including all of the specifically unobserved ones) is black? But I think that we are already familiar with such relations between facts about wholes and facts about parts. From observed events concerning visible bodies in chemistry, for example, we can infer many properties of their constituent atoms, which cannot be seen as individuals. It might, of course, be protested that the relevant micro-facts are indeed observable, though indirectly, precisely through their effects on larger bodies. But I could happily adopt the same formulation, and claim that inductive predictions are, after all, just another fallible form of indirect perception. We perceive, albeit dimly, that all ravens are black, and infer or indirectly seem to see (why should it matter which we say?) that each 'part' of all ravens, i. e. each individual raven, is black as well.

Consider this brief discussion:

Amy: How is the pizza at Mario's?

Bob: Pretty good. I've eaten there twice.

There are two ways to analyze Bob's epistemic situation here. One analysis is to say that Bob has tasted certain particular *slices* of pizza at Mario's on a certain two occasions, enjoyed them, and is now reporting an inductive inference to the effect that most of the millions of other slices of pizza at Mario's are equally good. The other analysis is to say that Bob has on two occasions tasted a certain general thing, namely *the pizza at Mario's*, found it pretty good, and is now reporting this directly as an observation. On the first analysis, Bob makes a thorough observation of a few entire small things (give or take some crumbs), about which he is able to judge with a high degree of certainty: those slices were pretty good. As to the pizza at Mario's generally, that should be seen as the set of all such slices, of which Bob has only tasted a tiny sample. Therefore, he is able to make only a fairly weak induction from his few samples to an enormous class, though such inductions are

supported by other inductive beliefs about the usual consistency of restaurant food.²⁰ On the second analysis, Bob has no greater total certainty about the general quality of Mario's pizza, since his observation of the stuff is slight and could easily be defeated by further experience. But he does, at least, have epistemic contact with the stuff as stuff, not just with members of a set. As a practical matter, it makes no difference which analysis we choose—although I think, as I have said above, that there is no good reason always to favor the first. What makes a difference here is that the first analysis leaves us with the problem of induction, while the second one does not. There is still the problem of grounding beliefs in sufficiently good evidence, and there is still the background problem of perception: how do we know that *any* observation is reliable? But there is no problem of induction where there is no induction.

But, is there really no induction here at all, or am I sneaking it in somehow? You must suspect that I am sneaking it in somehow. Based on my observations of some things, I am claiming to derive beliefs about other things that I have definitely not observed, for example ravens in the year 2050. What else can there be to connect the observed facts with the unobserved facts, other than some form of induction?

Here is my answer. There is indeed an inference from observed facts to unobserved facts, but it is a deductive, not an inductive inference. I observe the universal fact that *all* ravens are black, if it is a fact, when I observe the general fact that *ravens* are black, which I do at the same time that I observe the particular fact that *some* ravens are black. My belief that future ravens will be black is logically entailed by my belief that all ravens are black. It is not observed directly, but it does not have to be. There is no general law, after all, that the deductive consequences of our observational beliefs must be observed themselves, or even observable. Suppose a car goes by, and I observe that it is blue. I already know that all cars have registration forms, and that the color of each car is listed on its form. Therefore, I come to believe, based on my observation of this car, that the word 'blue' appears on its registration form, though I will have no opportunity to see the form itself. Now, it may be that this deductive inference yields a false conclusion, of the sort that everyone agrees induction sometimes produces. If I had made a

²⁰ At another restaurant, someone makes a little joke:

Carla: How is your filet mignon?

Dexter: I don't know yet. I've only eaten half of it.

This is a joke because we do commonly take our direct knowledge of parts and surfaces of most small things to count implicitly as knowledge of the whole things.

faulty observation of the car that went by, and it was really green instead of blue, then it would not say 'blue' on the car's registration form, so my deduced belief would be a false one. Similarly, if I falsely observe that all swans are white, based on my observations of swans in America, and deduce that swans in Australia are all white as well, then I am simply wrong. But what is wrong is not a faulty inference—my deduction was perfectly valid—just a misleading observation.

This proposed solution can be seen as providing an element that has always been missing from the classical hypothetico-deductive approach to scientific reasoning. On the hypothetico-deductive model, there is no such thing as an inductive argument *per se*. What happens instead, freely translated, is that scientifically interesting u-statements are initially written down only in pencil – that is, as mere hypotheses, not to be believed (because there is no initial reason to believe them), but just to be considered. Once they are on the list, we test them by deducing predictive p-statements from them, and then observing whether or not the predictions come true. In a standard version like Carl Hempel's, a hypothesis is held to be more believable the more it is confirmed by true predictions.²¹ In Popper's deductivist alternative, the hypothesis is never confirmed, but merely 'corroborated' by surviving attempts to find predictions that turn out to be false.²² Now, these procedures (one or both) strike most of us as a better description of actual scientific reasoning than simple inductive arguments. It does seem right to say that u-statements acquire greater credibility as they pass successfully through more comprehensive and more rigorous tests. But, as Salmon and others have pointed out, neither variant of the hypothetico-deductive approach provides a real solution to the problem of induction, because each fails to show how testing actually justifies belief in a hypothesis.²³ No account is given as to why one hypothesis should be initially considered rather than another, and it is not made clear why confirmation or corroboration makes the hypothesis in question more likely to be true than its surviving competitors.

²¹ Carl Hempel, *Philosophy of Natural Science* (Englewood Cliffs: Prentice-Hall, 1966).

²² Popper insists in *The Logic of Scientific Discovery* that he is not attempting to justify either induction or the hypothetico-deductive model, as he understands these terms. Instead, he wants his approach to be seen as entirely deductive.

²³ As Salmon points out in *The Foundations of Scientific Inference* (Pittsburgh: University of Pittsburgh Press, 1967), 25-26, if corroboration is supposed to give us any *reason to believe* the general hypothesis in question, based ultimately only on particular results of observations, then this amounts to an ampliative (hence non-deductive) element in Popper's theory, whatever he chooses to call it.

On the view that I am suggesting, however, our initial choice of one hypothesis over another can be accounted for, since some general statements will appropriately articulate our general observations, and some will not. An account can also be given of why both confirmation and non-falsification tend to add epistemic weight to these hypotheses. If we take the u-statement in question initially as the tentative report of an imperfect observation, then what are usually considered to be separate observations of confirming or non-falsifying instances can be seen instead as extensions and clarifications of the *same* observation. It would be a matter of making sure that our initial observation is a good one – in the same way that someone who thought he had seen an individual black raven might catch the bird and study it carefully, in order to add ink to his initial penciled-in report.²⁴

As long as there are some observationally acquired u-statements available from which appropriate theoretical hypotheses could be deduced, there is no need to hold that all types of general fact can be observed directly. It is in principle only necessary that there be *one* sufficiently general u-statement, the truth of which can be affirmed provisionally through observation – perhaps even something like “inductive inferences are generally reliable.” Kant tried to show that some such principle of nature's uniformity is knowable *a priori*, though Hume's arguments against that possibility seems to have proven more persuasive over time. In any case, once we had such a universal hypothesis penciled-in through observation, more specific u-statements could be deduced from it, and jotted down as likely to be true. The two-stage argument would go something like this:

(U1) Induction is reliable, i.e. if all observed A's are B, then probably, all A's are B (observed).

(U2) Therefore, if all observed ravens are black, then probably, all ravens are black (deduced from (U1)).

(U3) All observed ravens are black (observed).

(U4) Probably, all ravens are black (deduced from (U2) and (U3)).

²⁴ This is why we take some scientific experiments to yield general knowledge on the first try, and view repetitions as providing reassurance to our initial results, rather than new, logically separate facts. For example, it required only one carefully observed solar eclipse (in 1919) for physicists to perceive that light bends around massive objects.

In this way, the idea of inductive inference is ultimately vindicated by means of observation. But it is not vindicated as a fundamental form of reasoning – only as a certain conditional formula that has been observed to work well in general. The high-level principle of uniformity would not have to be observed in an immediate way, either. We could start with a few lower-level observations, to the effect that all ravens are black, all rats have tails, and the like. We could then submit some of these basic statements to the usual sorts of testing. If successful, the whole resulting situation could be said to be contained in an observation of the fact that this observational-deductive method usually works. Thereafter, we could with greater and greater confidence deduce unobserved hypotheses from the initially-weakly-observed general principle, and then through usually-successful testing add credence to both. This kind of ‘bootstrap’ procedure would require only that there be enough initial observational input at some level for the whole process to get going.²⁵

IV. Conclusion

In this paper, I have argued that the problem of induction, as it is usually conceived, presupposes the impossibility of our observing general facts. This is

²⁵ The idea of general observations might also help a bit with Nelson Goodman’s variant problem of induction. Goodman asks, in *Fact, Fiction, and Forecast*, 4th edition (Cambridge: Harvard University Press, 1983), how we can rationally choose to generalize on the basis of our ordinary concepts, like the colors blue and green, rather than such odd but clearly describable properties as ‘grue,’ which he defines as either green if first examined before a certain time *t*, or blue if first examined afterwards. Any prediction which ‘projects’ the property green before time *t* will be justified by precisely the same evidence, he says, as the corresponding prediction which projects the property grue—but clearly these are different, incompatible predictions, and arbitrarily many such equally incompatible, but equally well-evidenced, predictions could be generated just as easily. What good reason is there to prefer one to the other? One possible response is to replace the question of good reason with a question as to what one actually sees. If it is a fact that all emeralds are green, then this is one of the facts that one observes, when one observes a number of green emeralds in the absence of any observations to the contrary. But since it is not an actual fact that emeralds are grue, it is not, *a fortiori*, an observable fact. I could believe, of course, that I was seeing something grue, not green, when I was looking at a pile of emeralds. And this would certainly be a mistake, like looking at a solid green stone and somehow believing it was blue on the reverse side. But there may be some advantage to analyzing this mistake as a faulty observation, as distinct from an irrational inference. Grue, as defined, would not then be seen as an unprojectable property in principle, but rather as an invisible (or indistinguishable) one in practice. This is only to suggest an angle on Goodman’s problem, of course, not to pretend to have solved it.

why we seem to need inductive inferences to justify our general beliefs. But such inferences are hard to specify and seemingly impossible to justify in their own right; hence, the problem. I have attempted to undermine the common view of observation as always particular in scope, by arguing that the foundationalism, atomism, and individualism on which it seems to depend are all rightfully obsolete. I have suggested an alternative, holistic account of observation as a replacement, according to which general statements are indeed observable, albeit typically with low initial certainty. And I have tried to show how these defeasible general observations would neatly fit into the standard hypothetico-deductive model of scientific reasoning, by providing hypotheses, previously viewed as unempirical, with some measure of *prima facie* justification.

Somehow, I doubt that every reader has been totally convinced by these remarks to abandon the traditional idea of observation as exclusively particular, and to accept my sketch of a holistic account as adequate to the resolution of the problem of induction. But perhaps some readers are convinced to this extent: that the problem of induction depends on a certain theory of observation, that this theory is questionable, that a different theory can provide at least a superficial answer to the problem, and that there is an approach here worth exploring further.

SELF-EVIDENCE

Carl GINET

ABSTRACT: This paper develops an account of what it is for a proposition to be self-evident to someone, based on the idea that certain propositions are such that to fully understand them is to believe them. It argues that when a proposition *p* is self-evident to one, one has non-inferential a priori justification for believing that *p* and, a welcome feature, a justification that does not involve exercising any special sort of intuitive faculty; if, in addition, it is true that *p* and there exists no reason to believe that the proposition that *p* is incoherent, then one knows a priori that *p*. The paper argues that certain deeply contingent truths, e.g., the truth that I would now express by saying “I exist”, can be self-evident to, and thus known a priori by, the person they are about at the time they are about; but, since they cannot be known a priori, or even expressed, by anyone else or at any other time, they should not count as a priori truths.

KEYWORDS: self-evidence, a priori, justification,
non-inferential, knowledge

The term ‘a priori’ is much used by philosophers but there is not much agreement about what it means. There is wide agreement that it primarily refers to a way of knowing that is in some sense independent of experience (though in what sense is not agreed) and that certain sorts of truths (e.g., elementary truths of arithmetic) are ones we typically know a priori and certain other sorts of truths (e.g., truths as to what the weather was like yesterday) are ones we do not and cannot know a priori. A derivative use of ‘a priori’ is to specify a kind of truth, that which can be known a priori: a truth is a priori only if it is, or could be, known a priori. (Later I will raise a doubt as to whether that ‘only if’ should be strengthened to ‘if and only if.’) But beyond these points there is little agreement and I suspect that many who freely speak of the a priori have no well worked out view on the question of just what constitutes the a priori way of knowing.

I will try here to give an account of one basic sort of *a priori* knowing. It is a sort that I hope might serve as the foundation and key to a full account of the ways of knowing that should count as a priori; but that large (one might say grandiose) project I will not be able even to sketch here.

My account will assume without argument two things about knowing in general and one further thing about a priori knowing in particular. About

knowing in general I will assume, first, that one can know only facts, i.e., that if one knows that *p* then it is true that *p*; and second, that knowledge implies belief, i.e., that if one knows that *p* then one believes that *p*. About *a priori* knowing in particular I will assume that what makes a case of knowing that *p* a case of knowing *a priori* that *p* is the way in which the subject's belief that *p* is justified: to say that a piece of knowledge is *a priori* is to say that the belief involved is justified *a priori*.

Belief justifications in general, whether *a priori* or not, divide exhaustively and exclusively into two kinds, inferential and non-inferential. In this paper I will try to explicate a non-inferential kind of *a priori* justification. The key notion in my explication will be *self-evidence*. I aim for an account such that if it is self-evident to a person that *p* then that person has a *a priori* and non-inferential justification for believing that *p*; and if it is also true that *p* then, normally, they know that *p* and its being self-evident to them will be their *way of knowing* that *p*. And I want an account on which its being self-evident to a person that *p* will *not* be a matter of their exercising any special sort of intuitive faculty. Later I will say something about why I take this to be an important desideratum.

The Account of Self-Evidence

I believe I can give an account that has these desirable qualities by exploiting the idea (to put it in a simple preliminary way) that some propositions are such that fully understanding them requires believing them. Believing them is at least part of what it is to fully understand them; they state a condition such that belief that the condition holds is constitutive of grasping the proposition.¹

¹ This idea is, of course, not new. In W.V. Quine and J.S. Ullian, *The Web of Belief* (New York: Random House, 1970) statements are said to be self-evident just in case "to understand them is to believe them." Christopher Peacocke use the notion of a belief's being constitutive of the possession of a concept in giving an account of *a priori* truth; understanding a proposition requires, of course, understanding the concepts involved in it (See his "How Are *A Priori* Truths Possible?," *European Journal of Philosophy* 1, 2 (1993), 175-199, and "Implicit Conceptions, the *A Priori*, and the Identity of Concepts," in *Concepts*, ed. Enrico Villanueva (Atascadero: Ridgeview, 1998), 121-148). George Bealer, in explaining how some intuitions are *a priori* sources of evidence, uses the idea that for some propositions 'determinate' understanding of one of them requires an intuition that it is true (but for Bealer intuitions are not beliefs and it is not clear whether, on his view, if an intuition that *p* is a source of evidence, a belief that *p* is justified; also his account of why such intuitions are evidence relies on the idea that they reliably indicate truth, whereas my argument that one's belief that *p* is justified when it is self-evident to one that *p* does not assume a correlation between self-evidence and truth) (see his "A Theory of the *A Priori*," *Philosophical Perspectives. Epistemology* 13 (1999): 29-55).

I shall first present two definitions, one of what it is for a proposition to be self-evident (period) and the other of what it is for a proposition to be self-evident to a particular person. Then I will go on to argue that when a person fully understands and believes a proposition that is self-evident to them they are justified in believing it. Here is a preliminary version of the first definition, preliminary because it will need a couple of qualifications, to be explained later (in this definition and hereafter I will use the phrase “what the sentence **p** says” as short for the phrase “what is said by one who utters **p** in normal circumstances in order to assert that **p**”):

(D1-prelim-1) For any declarative sentence **p** whose meaning is such that what the sentence **p** says does not vary from one context of utterance to another, it is self-evident that **p** if and only if: the sentence **p** is such that, for any person *S*, if *S* understands what the sentence **p** says then it follows that *S* believes that **p**, expressed that way.²

By ‘*S* believes that **p**, *expressed that way*’ I mean that *S* knows that the sentence **p** (and not just some other sentence that says the same thing) expresses what he believes: either he believes that what **p** says is true or, in suitable circumstances, he would sincerely assert the sentence **p** or would sincerely assent to another’s assertive use of that sentence. (Note that the restriction that what the sentence says does not vary from one context to another means that this definition does not apply to any sentence containing indexical terms, such as the sentence “I exist”. Later I will consider whether (and how) what such a sentence says can be said to be self-evident to its utterer.)

Relativity to Sentences

The definiens of (D1-prelim-1) speaks of what the sentence **p** says, rather than the proposition that **p**, partly because the former is what I mean by talk of the proposition that **p** and I might as well say what I mean. Also, I find it easier to know what I’m talking about when I talk about understanding what the sentence **p** says than when I talk about understanding the proposition that **p**. Understanding what the sentence **p** says can be explained in terms of understanding the parts and the structure of the sentence; there is no comparably

² I use the plaintext letters ‘*p*’ and ‘*q*’ as variables ranging over sentences and the boldface versions as variables ranging over names of sentences. Thus a value for an occurrence of ‘*p*’ would be a sentence and a value for an occurrence of ‘**p**’ would be a sentence in quotes.

clear and unproblematic way of explaining what is involved in understanding the proposition that *p*.³

But speaking of understanding what a sentence says, rather than of understanding a proposition, introduces into the object of understanding a relativity to sentences. It makes the particular sentence mentioned a part of what is understood (namely, what that sentence says). For it may be that someone understands what sentence *p* says but not what a different sentence *q* says, even though what *p* says is the same as what *q* says, i.e., they express the same proposition.

This relativizing of understanding to sentences might be thought to introduce a problem for our definition of self-evidence. If there are sentences *p* and *q* that say the same thing but are such that *p* satisfies the definiens of (D1-prelim-1) but *q* does not, then our definition would force us to say that what *p* says is self-evident but what *q* says is not self-evident, even though what *p* says is the same as what *q* say—sa violation of Leibniz's law.⁴

³ I do not mean to suggest, however, that understanding a proposition must always amount to understanding what some sentence says. It seems clear that it need not. It seems possible, for example, that there should be a person who does not understand the sentence 'Any triangle has three internal angles' or any other sentence that says the same thing, but who does understand the proposition that a triangle has three internal angles. Such a person could have formed the abstract conception of a triangle (by visual abstraction from various seen triangles) without giving it a name, and similarly for the notion of an internal angle, and to could see, by a kind of visual abstraction, that having three internal angles is a part of that conception. This would be to understand the proposition that a triangle has three internal angles *graphically*, so to speak, rather than via a linguistic medium.

⁴ One might be tempted to think that there is a more conclusive argument than the one I give below for the conclusion that there cannot be sentences *p* and *q* that say the same thing and such that one satisfies (D1-prelim) and the other does not. This conclusion would follow from the lemma that there cannot be two sentences saying the same thing that are such that someone understands what the one says but not what the other says. And one might think that this lemma follows, by Leibniz's law, from the premises (1) *S* understands what *p* says and (2) what *p* says is identical with what *q* says. But it does not. Premise (1) does not provide an extensional context for the referring term 'what *p* says,' such that the truth-value of (1) must be preserved by substitution of any co-referring term. This is because (1) entails that (3) *S* knows some truth of the form 'What *p* says is that *r*.' And from (3) and (2) it clearly does *not* follow that (4) *S* knows a truth of the form 'What *q* says is that *r*.' Note that (1), the proposition that *S* understands what *p* says, differs from the proposition that *S* *believes* what *p* says. The latter does provide an extensional context for 'what *p* says': if what *p* says is the same as what *q* says, then the proposition that *S* believes what *p* says does entail that *S* believes what *q* says.

But is there any pair of sentences that fills this bill? The most plausible sort of candidate I know of⁵ can be illustrated as follows. Consider the sentence (in decimal notation):

$$(A) \quad 10 + 10 = 20$$

That seems a good candidate for being a sentence such that you could not count as understanding what it says if you didn't believe what it says. Now consider the following sentence in binary notation:

$$(B) \quad 1010 + 1010 = 10100$$

It certainly seems possible that someone who understands binary notation well enough to understand what (B) says might not yet see that what (B) says is true (especially if they were able to understand what (B) says without 'translating' it into decimal notation, i.e., into (A)).⁶ If so, I would want to say that it is self-evident that (speaking in decimal notation) $10+10 = 20$, but it is not self-evident (speaking in binary notation) that $1010 + 1010 = 10100$.

What (B) says might, however, seem to be the very same thing as what (A) says. After all, '1010' in binary notation designates the same number as '10' designates in decimal notation, and '10100' in binary notation designates the same number as '20' designates in decimal notation. But, I hope to persuade you, what (A) says and what (B) says are *not* the same thing.

What *does* each of the sentences in question say? (A) is in decimal notation. If you know how to interpret decimal notation you know that, for any numeral in that notation, $\mathbf{d_n} \dots \mathbf{d_2 d_1}$, the first digit on the right, $\mathbf{d_1}$, gives the number of ones [$\mathbf{d_1} \times 10^0$], the second digit from the right, $\mathbf{d_2}$, the number of tens [$\mathbf{d_2} \times 10^1$], the third digit, $\mathbf{d_3}$, the number of hundreds [$\mathbf{d_3} \times 10^2$], and so on.⁷ So a decimal representation consisting of a single digit $\mathbf{d_1}$ is read as naming the number that is $\mathbf{d_1} \times 1$, one of two digits $\mathbf{d_2 d_1}$ means $(\mathbf{d_2} \times 10) + (\mathbf{d_1} \times 1)$; one of three digits $\mathbf{d_3 d_2 d_1}$ means $(\mathbf{d_3} \times 100) + (\mathbf{d_2} \times 10) + (\mathbf{d_1} \times 1)$; and so on. So what sentence (A) says to one

⁵ Suggested to me by Bob Stalnaker in conversation many years ago.

⁶ I use scare quotes around 'translating' because, as I go on to argue, (A) is not a translation of (B) in the sense that it says the same thing.

⁷ I use the boldface letters with subscripts ' $\mathbf{d_1}$ ', ' $\mathbf{d_2}$ ', etc. as variables ranging over the single-digit numerals '0,' ..., '9'—i.e., names of the numbers 0, ..., 9—and I use the plaintext versions as variables ranging over those numbers.

who takes it to be in decimal notation and understands decimal notation is the following:

(A*) One ten plus one ten equals two tens.

And that is surely self-evident, as self-evident as what is said by any sentence of the form ‘One X plus one X equals two Xs’, as self-evident as that $1+1=2$.

Sentence (B) is in binary notation. If you know how to interpret binary notation you know that the first digit on the right, **d1**, gives the number of ones [$d1 \times 2^0$], the second digit from the right, **d2**, the number of twos [$d2 \times 2^1$], the third digit, **d3**, the number of fours [$d3 \times 2^2$], the fourth digit, **d4**, the number of eights [$d4 \times 2^3$], and so on. So a binary representation of a single digit **d1** is read as naming the number that is $d1 \times 1$; one of two digits **d2d1** means $(d2 \times 2) + (d1 \times 1)$; one of three digits **d3d2d1** means $(d3 \times 4) + (d2 \times 2) + (d1 \times 1)$; one of four digits **d4d3d2d1** means $(d4 \times 8) + (d3 \times 4) + (d2 \times 2) + (d1 \times 1)$; and so on. So what sentence (B) says to one who takes it to be in binary notation and understands binary notation is the following:

(B*) One eight plus one two, plus one eight plus one two, equals one sixteen plus one four.

In decimal notation it says that $(8+2) + (8+2) = 16+4$. And that is *not* self-evident. A relative beginner at arithmetic could fail to see that what “ $(8+2) + (8+2) = 16+4$ ” says is true, even though she fully understands what it says—as evidenced, say, by her knowing how to go about figuring out whether it is true.

But it’s also pretty intuitive that what (B*) says is *not* what (A*) says. These arithmetical truths are clearly *different* truths. The arithmetical fact that $10 + 10 = 20$ is not the same as the fact that $(8+2) + (8+2) = 16+4$, and both are different from, for example, the fact that $(2^2+2^2+2) + (2^2+2^2+2) = 5^2-5$ —despite its being the case that “10”, “ $8+2$ ”, and “ 2^2+2^2+2 ” all name the same number, as do “20”, “ $16+4$ ”, and “ 5^2-5 ”. Someone might know one of these facts while not yet knowing the others.

It is indeed compatible with fully understanding what (A) and (B) say that one believe what (A) says and not what (B) says, but this is because they do not say the same thing. In general, if two sentences do say the same thing and what they say is self-evident⁸, then anyone who fully understands both sentences must

⁸ Without this second conjunct in the condition, Kripke’s well-known example of Pierre in London/Londres (Saul Kripke, “A Puzzle About Belief,” in *Meaning and Use*, ed. Avishai

believe what each of them says.^{9,10} So pairs of sentences like (A) and (B) do not after all present a problem for definition (D1-prelim-1).

Examples

Here are some examples of sentences whose meaning in English is such that, according to (D1-prelim-1), what they would say is self-evident.

- (1) (a) One plus one is two.
- (b) If there are a man and a woman and the man is the woman's brother, then the woman is the man's sister.
- (c) Any triangle has three internal angles.
- (d) For any true or false propositions p and q : if it is true that $p \leftrightarrow q$ and true that p , then it is true that q .
- (e) A rhombus is a Euclidean plane figure.
- (f) A rhombus is a Euclidean closed plane figure with four equal sides.
- (g) If there are two lines, A and B , and line A is longer than line B , then line B is shorter than line A .

To say that what sentence (1)(a) ("One plus one is two") says is self-evident is, according to (D1-prelim-1), to say that the meaning of the sentence is such that if a person fully understands what it says then it follows that they believe what it says, namely, that one plus one is two. The idea, to put it contrapositively, is that if a person hesitates to accept what 'One plus one is two' says, is uncertain whether one plus one is two, then it follows that the person does not fully understand what that sentence says. Believing that one plus one is two is a

Margalit (Dordrecht: Reidel, 1979)) would, as comments from Juan Comesana led me to see, provide a counter-example to the claim.

⁹ Obviously, in my view to individuate the things sentences say, propositions, in such a way that they are identical if necessarily equivalent is not to individuate them finely enough.

¹⁰ Can we make the further claim that, if two sentences say the same self-evident thing, then anyone who fully understands both sentences must see that they say the same thing? No. Kripke's example of Pierre in London/Londres provides a counter-example. Consider the sentences "If London exists and is pretty then London exists and is pretty" and "Si Londres existe et est jolie alors Londres existe et est jolie." These say the same thing, a thing that is self-evident by my definition. Yet someone who mistakenly thinks that 'London' and 'Londres' name different cities will mistakenly think that those sentences say different things; and yet such a person might know what city 'London' names and also what city 'Londres' names, by, e.g., having been directly acquainted with the city under that name.

necessary, constitutive condition of fully understanding what “One plus one is two” says.

By way of contrast, here are some sentences such that what they say is not self-evident:

- (2) (a) Seventy four times twenty three is one thousand seven hundred and two.
- (b) Most men with sisters are taller than they are.
- (c) The sum of the internal angles of any triangle equals the sum of two right angles.
- (d) For any true or false propositions p , q , and r :
it is true that $[p \leftrightarrow (q \leftrightarrow r)] \leftrightarrow [(p \leftrightarrow q) \leftrightarrow r]$.
- (e) Rhombuses shown in geometry textbooks are often not square.

To say that what sentence (2)(c) (“The sum of the internal angles of any triangle equals the sum of two right angles”) says is not self-evident is to say that it is possible for someone to fully understand what that sentence says and yet fail to accept what it says, be uncertain whether the sum of the internal angles of any triangle equals the sum of two right angles. That a person hesitates to accept what (2)(c) says does not entail that the person fails to fully understand what (2)(c) says.

Full Understanding

Here there might arise an objection, one that presses me to say more about what constitutes *full* understanding. Why shouldn’t we take a person’s failure to accept what (2)(c) says (that the sum of the internal angles of any triangle equals the sum of two right angles) as a failure to *completely* understand what it says? In particular, as a failure to completely understand what a triangle is? Doesn’t coming to learn the truth of what (2)(c) says, by seeing a proof of it, make one’s understanding of the essential nature of triangles more complete? And shouldn’t someone who lacks this more complete understanding of what a triangle is be said to lack a *full* understanding of what is said by sentence (2)(c)?

Analogously, this questioner might say, someone may have a partial understanding of what a rhombus is, enough to know that a rhombus is a Euclidean plane figure, but lack the full understanding that would entail believing what (1)(f) says (that a rhombus is a Euclidean plane figure with four equal sides). If fully understanding what (1)(f) says entails believing it, why doesn’t fully understanding what (2)(c) says entail believing it? If I want to say (as I do) that fully understanding what (1)(f) says entails believing it but that fully understanding what (2)(c) says does not entail believing it, then I need to give a

plausible account of *fully* understanding what a sentence says that will yield this discrimination between the two cases.

There is, of course, a sense in which one who knows the fact expressed by (2)(c), that the sum of the internal angles of any triangle equals the sum of two right angles, understands the nature of triangles better or more completely than one who does not know this fact. But it does not follow that the latter person cannot be one who fully understands what sentence (2)(c) says. For what is required to fully understand what that sentence says is just that (a) one grasps the concept expressed by each of its descriptive (contentful) terms—e.g., ‘triangle,’ ‘internal angles,’ ‘sum,’ ‘equals,’ ‘right angles,’ ‘two’—well enough to be able to tell with respect to any candidate case, given sufficient relevant information about it, whether the concept applies in that case—we can speak of this as having *application-competence* with respect to the term—and (b) one correctly perceives the grammar of the sentence, i.e., one understands the way the sentence is put together well enough to know how the meaning of each of its descriptive terms contributes to what the sentence says. And certainly one who is ignorant of the truth of what (2)(c) says may nevertheless satisfy these requirements with respect to (2)(c). He may have application competence with respect to every one of (2)(c)’s descriptive terms and grasp its grammar perfectly and yet fail to believe what it says.

In contrast, one who is uncertain whether to accept what (1)(f) says (that a rhombus is a Euclidean plane figure with four equal sides) fails to have application-competence for the term ‘rhombus’ (assuming that they have application-competence for the other descriptive terms in (1)(f) and grasp the grammar of the sentence): there will be particular Euclidean plane figures such that they will be unable to tell whether those figures are rhombuses no matter how much relevant information they have about them (of a sort that can be acquired without already having the concept of a rhombus). They may have partial application competence for the term ‘rhombus,’ since application competence can come in degrees. They may know that ‘rhombus’ denotes a Euclidean plane figure—i.e., that what sentence (1)(e) says is true—but nothing about what sort of plane figure, or they may know that it designates one with four sides but nothing more. If their application competence for ‘rhombus’ falls short of full competence then their understanding of what (1)(f) says falls short of full understanding.¹¹

¹¹ Partial application competence can come in different varieties with different sorts of terms. Consider color terms. A young child might happily apply the term ‘red’ to that shade we call

Similarly for the other sentences in (1), e.g., (1)(c). One who does not believe what sentence (1)(c) says (that any triangle has three internal angles) must fail to fully understand what it says, i.e., must lack application-competence for one or more of its descriptive terms ('triangle,' 'three,' 'internal angles'), or fail to grasp fully the grammar of the sentence, or both. No other explanation for the failure to accept what the sentence says is possible.

(I should note that, though adequate for present purposes, my explanation of full understanding is incomplete. I have described application-competence for the descriptive terms in (1)(c), (1)(e), (1)(f), (2)(c), and (2)(f)—terms such as 'triangle,' 'rhombus,' 'right angle,' 'Euclidean plane figure,' 'three'¹²—as requiring the ability to determine with respect to any candidate case, given sufficient relevant information about it, whether the term applies in that case. Such a requirement is apt for those terms and many others in mathematics and logic, but it will not be apt for many other descriptive terms—for example, terms that are vague ('bald,' 'red,' 'tall'), evaluative terms whose meaning makes their application essentially contestable ('expensive'), and terms denoting natural kinds about which there are necessary truths that are only empirically discoverable ('water,' 'elm,' 'tiger')—and it is not apt for proper names or other directly referential terms ('Hannah,' 'London,' 'that man over there'). For terms of these sorts it will be necessary to complicate in one way or another the specification of what is required for application competence and thus what is required for full understanding of sentences in which such terms occur. I will not try here to work out these complications. For a great many of the sentences containing such terms that say things that are self-evident, it will be clear that their doing so does not depend on what the right account of those complications is—such as, e.g., "A man with no hairs on his head is bald," "An expensive car is not a cheap car," "Liquid water is wet," "Elms are not animals," "If there are vixens, then all vixens are female foxes,"¹³ "Provided that Sarah and Hannah exist, if Sarah was born earlier than Hannah was born, then Hannah was born later than Sarah was born."^{14, 15})

'fire-engine red' but hesitate or refuse to apply it to other, more orangeish or bluish, shades of red.

¹² Application competence for 'triangle,' 'rhombus,' and 'right angle' will require being able to tell, given enough relevant information, whether something is a triangle, a rhombus, or a right angle. Application competence for the name of a number, like 'three,' will require knowing where it comes in the counting order, which means knowing how to list the number names (some version of them) in the right order, at least up to the number name in question.

¹³ Someone who fully understands the shorter sentence "All vixens are female foxes" might not accept it as saying something true if they doubt the existential presupposition that there are

Qualifications

I said above that for sentences like those in my list (1), if a person doubts what is said by one of them, no explanation of this is possible other than that they fail to fully understand what is said by the sentence. Actually, two other sorts of explanation may sometimes be possible for some sentences such that what they say is self-evident.

The first sort of explanation I have in mind is a *recherché* possibility and clear illustrations of it are hard to come by, but the following may suffice to explain the idea. Consider the sentence

(3) For any property actually possessed by some entities, there is a set whose only members are all of the things that possess that property.

It's plausible to think that there was a time when a person's failure to believe what this sentence says would have been knockdown evidence that they lack application-competence for the term 'set,' that they fail to fully understand what a set is (assuming they have application-competence for the other descriptive terms in the sentence and grasp the grammar of the sentence). (If this is not an actual historical truth, it represents a possible history.) Then came Russell's paradox: a deduction of a contradiction from what this sentence says and things said by other sentences that they took to be (and we still take to be) necessary truths (namely, that the property of being an entity that does not have the set-membership relation to itself is a genuine property [indeed, one possessed by many actual things] and that everything must either possess or lack that property). That proof

vixens, but they would have to believe what is said by the self-evident "If there are vixens, then all vixens are female foxes."

¹⁴ Someone who fully understands the shorter sentence "If Sarah was born earlier than Hannah was born, then Hannah was born later than Sarah was born" might not regard it as saying something true if they were not sure that the names 'Sarah' and 'Hannah' both refer, but they would have to believe what is said by the self-evident "Provided that Sarah and Hannah exist, if Sarah was born earlier than Hannah was born, then Hannah was born later than Sarah was born."

¹⁵ Proper names offer a plausible example for which the decision as to whether a sentence containing them says something self-evident does depend on the right account of application competence for that sort of term. Consider sentences of the form "If *a* exists, then *a* = *a*" where *a* is a proper name. If the convention in the language in which such a sentence occurs is that occurrences of the same name within the same sentence must refer to the same object (as in many formal languages), then what such a sentence says is self-evident. But if there is no such convention (as arguably there is not in English), then what it says is not self-evident.

gave people who fully understood what sentence (3) says reason to believe that what it says is *incoherent*—in the sense that its negation can be deduced from necessarily true premises—and thus reason to doubt, indeed deny, that what it says is so. And this gave them reason to think that the notion of a set, as hitherto defined by the axioms of ‘naïve’ set theory (as we now call it), was incoherent.

This suggests that, to cover the sort of circumstance illustrated in the example, we need to amend our definition of self-evidence. Although that sort of circumstance does not obtain with respect to what is said by any of the sentences in (1) and is highly unlikely ever to obtain, it seems that it is possible in some sense that it should do so. Thus, if we want to guarantee that our definition of self-evidence covers what is said by those sentences, we need to revise it by inserting a disjunct to allow for this possibility.

So revised our definition will look like this:

(D1-prelim-2) For any declarative sentence **p** whose meaning is such that what the sentence **p** says does not vary from one context of utterance to another, it is self-evident that **p** if and only if: anyone who fully understands what they would say by uttering the sentence **p** and who does not have reason to believe that what **p** says is incoherent must believe that **p**, expressed that way.

Another qualification seems called for by some examples that Williamson¹⁶ has given in order to cast doubt on the sort of claim that I am making (the claim that there are, as Williamson puts it, “understanding-assent links”). He describes cases of people who fully understand sentences of forms like “Every F is an F” or “If P and Q, then P” yet do not assent to them because of their views about the logic of non-atomic sentences where some of the atomic components are neither true nor false. (For example, on some views, if for a particular value of *x* “*x* is a vixen” is neither true nor false then “if *x* is a vixen then *x* is a vixen” is likewise neither true nor false, and if any instance of that open conditional fails to be true then so does “If there are vixens, then every vixen is a vixen.”) Such examples won’t be possible for sentences for which there is not the possibility of their being neither true nor false and many sentences of mathematics and logic seem to be of that sort. However, even for sentences that can be neither true nor false, our definition of ‘self-evident’ can be amended to rule out the kinds of counterexamples Williamson describes, by adding to the clause ruling out intelligible reasons for declining to assent while fully understanding the sort of reason that Williamson describes. Thus we make that clause read “who does not

¹⁶ Timothy Williamson, *The Philosophy of Philosophy* (Malden: Blackwell, 2007), ch. 4.

have reason to believe that what **p** says is incoherent *and does not have reason to believe that it is neither true nor false.*"

I think that no further qualification is needed, for it does seem clear that we could not make sense of someone who fully understands "If there are vixens, then every vixen is a vixen" (and has no reason to believe that what it says is incoherent) and yet refuses to assent to what it says except by supposing that they have some such reason having to do with the logic of truth-valueless sentences. If an ordinary competent speaker who is not a philosopher or logician and is entirely innocent of such ideas (or of the idea that what the sentence says might be incoherent), if such a person hesitates to accept what that sentence says, then that is conclusive evidence that they do not fully understand what it says.

So I think we may state our final definition of 'self-evident' as follows:

(D1) For any declarative sentence **p** whose meaning is such that what the sentence **p** says does not vary from one context of utterance to another, it is self-evident that **p** if and only if: anyone who fully understands what they would say by uttering the sentence **p**, who does not have reason to believe that what **p** says is incoherent and does not have reason to believe that it is neither true nor false, must believe that **p**, expressed that way.

Self-Evidence is Noninferential A Priori Justification

I am now in a position to present my second definition, of what it is for a proposition to be self-evident *to* someone:

(D2) For any declarative sentence **p** whose meaning is such that what the sentence **p** says does not vary from one context of utterance to another, it is self-evident to S that **p**, expressed that way, if and only if: it is self-evident that **p**, **S** fully understands and believes that **p**, expressed that way, and **S** does not have reason to believe that what sentence **p** says is incoherent or reason to believe that what it says is neither true nor false.

(Hereafter, when I say things of the form "It is self-evident to **S** that **p**" without the qualification "expressed that way," I will mean that, for some sentence **q** which says the same thing as **p** says [**q** could, but need not, be the sentence **p** itself], it is self-evident to **S** that **q**, expressed that way.)

I claim that

(J) If it is self-evident to **S** that **p** then **S** is justified in believing that **p**.

Consider again sentence (1)(c) ("Any triangle has three internal angles"). Someone who does not believe what that sentence says—who denies that what it says is

true or is uncertain whether it is true—does not fully understand what that sentence says (provided that he does not have reason to believe that what the sentence says is incoherent or is neither true nor false). Someone who does fully understand what it says (where the just mentioned proviso holds: hereafter I will use the predicate ‘is innocent’ to mean that its subject satisfies this proviso), and therefore believes it, is, according to (J), justified in believing it.

How so? The fact that constitutes her being justified is simply the fact that she fully understands what the sentence says (and is innocent). This *entails* that she believes what it says. So it cannot be that she ought not to believe it *even though* she fully understands what it says (and is innocent). Nor can it be that she ought not to understand what it says (or that she ought not to be innocent). If she cannot be rationally criticized for fully understanding it (or for being innocent), then she cannot be criticized for what these things entail, her believing it. Therefore, given that she fully understands what it says (and is innocent), she is justified in believing it.

The fact that she fully understands it (and is innocent) *is* her justification, and this fact clearly does not constitute an inferential justification: it involves no belief in any premise such that she believes that what the sentence says is legitimately inferable from that premise. If what a sentence says is self-evident to a subject, then the subject’s belief in it is *non-inferentially* justified.

And this sort of justification is clearly a priori (if any is). It certainly satisfies any plausible negative constraint on a priori justification: it is not justification by sense perception or by introspection; nor is it by inference ultimately from perceptual or introspective beliefs. Indeed, the only experience that justification by self-evidence requires of its subject is whatever was needed in order to fully understand what the sentence in question says, the experience needed in order to know (a posteriori) some appropriate truth of the form “What *p* says is that *r*,” but that sort of dependence on experience is no reason to deny that the justification is a priori.

An important merit of justification by self-evidence, as I have explained it, is that it involves no appeal to any *special* way of coming to believe what the sentence says—by clear and distinct perception, or rational intuition, or the like. It involves just fully understanding what the sentence says (and being innocent) and therefore believing it. The justificatory force does not arise from any special justificatory quality attaching to the understanding or to the believing or to some accompanying mental state. It arises just from the fact that for these special sentences fully understanding what one of them says (while being innocent) entails believing it. This is important because our account of justification by self-

evidence, in being free of appeal to any special sort of mental state or process, is free of any of the difficulties or disputes that are apt to arise about the nature of such special states and about their credentials as justifiers. The account avoids having to face the question of what defines such a special state or process and the question of how its nature justifies one in believing its deliverances.

Does belief entail full understanding?

So in certain sorts of cases, fully understanding what a sentence says (while being innocent with respect to that sentence) entails believing what it says. Is there also an entailment the other way, from believing to fully understanding? Consider sentence (1)(e), “A rhombus is a Euclidean plane figure.” Must someone who believes what that sentence says, believes that a rhombus is a Euclidean plane figure, fully understand what it says? Well, no, it is not necessary that they fully understand that (or any other) English sentence. A monolingual speaker of Japanese can believe that a rhombus is a Euclidean plane figure. The question I really want to ask is this: If someone believes that a rhombus is a Euclidean plane figure, *expressed that way*, does it follow that he fully understands what the sentence “A rhombus is a Euclidean plane figure” says?

Suppose S knows that the term ‘rhombus’ designates a kind of Euclidean plane figure but does not know that a rhombus is a Euclidean plane figure with four equal sides whose opposite sides are parallel; S lacks application-competence for the term ‘rhombus’ and so does not fully understand what is said by the sentence “A rhombus is a Euclidean plane figure.” And suppose that S lacks application-competence for any synonymous term, in English or any other language; i.e., S does not fully understand any sentence that says that a rhombus is a Euclidean plane figure. Does it follow that, although S can believe that the word ‘rhombus’ designates a kind of Euclidean plane figure, S cannot believe that a rhombus is a Euclidean plane figure?

It is clear that at least some understanding of what a sentence **p** says, and more than a tiny bit, is necessary for being in a position to believe that **p**, expressed that way. I find in a textbook on topology¹⁷ after the word ‘THEOREM’ the following sentence:

If X is a locally compact topological space which is either Hausdorff or regular, then the family of closed compact neighborhoods of each point is a base for its neighborhood system.

¹⁷ John L. Kelley, *General Topology* (Princeton: Princeton University Press, 1955), 146.

I am quite confident, and with plenty of justification, that what the author says with that sentence is true. But if I were to say, “I believe that if X is a locally compact topological space ... etc.,” I would thereby pretend to a much greater comprehension of the concepts of topology than I in fact possess.

Suppose S knows very little English. S has learned that the word ‘rhombus’ designates some sub-kind of the kind of thing that “Euclidean plane figure” designates, but doesn’t know what kind either term designates, doesn’t have application-competence for either term. It would surely be misleading for S to say, “I know (believe) that a rhombus is a Euclidean plane figure.” S would imply that he understands what the sentence “A rhombus is a Euclidean plane figure” says more fully than he does.

How well does someone have to understand what a sentence says in order to be in a position to believe what it says? If we don’t require full understanding and don’t allow scant understanding, where in between shall we draw the line? I am unable to see any way of specifying a kind or degree of understanding other than full understanding that would clearly be enough for believing. And I’m inclined to think that, if we lack any principled and motivated way of drawing a line somewhere *between* scant and full understanding, then it would be arbitrary to draw it anywhere short of full understanding. Therefore, it should be drawn at full understanding. We should rule that a person cannot correctly say of herself “I believe that p ” if she does not fully understand what p says.

We should, however, allow that a person who does not fully understand what p says might convey a truth about herself by falsely saying “I believe that p .” She may convey at least the truth that she believes that what the sentence p says is true. And she may convey something more, about what she takes the relation(s) among the meanings of the words in the sentence to be. Someone who doesn’t know that a rhombus is a Euclidean plane figure with four equal sides may say “I believe that a rhombus is a Euclidean plane figure” and mean that she believes that the word ‘rhombus’ designates a kind of Euclidean plane figure. (Such a person, though failing to satisfy our criterion for full understanding of the sentence “A rhombus is a Euclidean plane figure,” because she lacks application-competence for “rhombus”, may nevertheless fully understand the sentence “The word ‘rhombus’ designates a kind of Euclidean plane figure.”)

If we have it that S ’s believing that p , expressed that way, entails S ’s fully understanding what p says, we can affirm the following: In a case where what p says is self-evident, S ’s believing that p , expressed that way, (while not having reason to believe that what p says is incoherent or lacks truth-value) entails S ’s being justified in believing that p . If S believes that p , expressed that way, then S

fully understands what **p** says; and if what **p** says is self-evident (and S lacks reason to think it is incoherent or without truth-value), then by the argument given earlier, S's belief that **p** is justified.

Self-Evidence and Knowledge

If S's belief that **p** is true, as well as justified by self-evidence, does it follow that S knows that **p**? Not necessarily. Suppose that, although the proposition that **p** is an elementary truth of logic or mathematics, it has recently become widely (but mistakenly) believed by experts that the proposition that **p** is incoherent, but this fact is unknown to S, a non-expert. Such a fact would, it seems to me, require us to judge that S does not know that **p**, despite having a true belief that **p** justified by self-evidence. It would be nice to be able to illustrate this intuition with a sentence **p** such that we'd be prepared to grant that (a) it is true that **p**, (b) there is someone to whom it is self-evident that **p**, but (c) it is believed by experts that the proposition that **p** is incoherent. But I can think of no such sentence.

My earlier example of sentence (3) ("For any property actually possessed by some entities, there is a set whose only members are all of the things that possess that property.") is not one about which we are prepared to grant that what it says is true. However, perhaps it can nevertheless be used to construct a fictional example that might serve as an illustration. Suppose that eventually some, but not all, set-theory experts come to think that what (3) says *is* true (that the best way to develop set-theory, after all, is to take that proposition as an axiom and reject some other premise involved in Russell's paradox). To them, however, what (3) says, though true, will not be self-evident, for they are still aware of some reason to think that what sentence (3) says is incoherent—namely the fact that Russell's paradox is still held by some experts to be a sound proof—even though they now have what they take to be better reason to think that Russell's proof is not sound and that what sentence (3) says is not after all incoherent. Suppose now that they consider Schmege, a logician of the late nineteenth-century before the discovery of Russell's paradox, to whom what (3) says was self-evident. Should they not judge that, although Schmege's belief in what (3) says was justified and true, it was not knowledge – precisely because, though Schmege did not realize it, a contradiction can be deduced from what (3) says and other propositions that were also self-evident to Schmege? I think they should.

If this is right, then S's having a true belief that **p** justified by self-evidence is not sufficient for S's knowing that **p**. It must also be the case that there does not exist reason to believe that the proposition that **p** is incoherent. The relation between self-evidence and knowledge is then as follows: if it is true that **p**, it is

self-evident to S that p, and there does not exist reason to believe that the proposition that p is incoherent, then S knows that p.

(K) S knows a priori that p if: it is true that p, it is self-evident to S that p, and there does not exist reason to believe that the proposition that p is incoherent.

This is an a priori way of knowing that p, because the justification involved, self-evidence, is a priori.

There are those who deny that a priori knowledge is possible.¹⁸ How are they likely to react to claim K? It would not, I think, be plausible to deny that if someone satisfied the condition laid down in K then they would know a priori that p. Deniers of a priori knowledge would, I think, instead claim that this condition cannot be satisfied, specifically, that no sentence does, or can, satisfy our definition of what it is for something to be self-evident to S (D2) because none can satisfy our definition of self-evidence (D1).

But what case could be made for this denial? It should, I think, strike one that there must be self-evident sentences (ones satisfying (D1)) once one notes that some beliefs must be part of understanding any sentence, even a false one. The difference between one who understands what a sentence says and one who does not (but is in other respects the same) must consist, at least partly, in beliefs had by the former that are not had by the latter. For example, the difference between one who fully understands what "All mice are blind" says and one who does not, might consist in the fact that the first, but not the second, believes that if something is blind then it cannot see. And the difference between one who fully understands "All flammable materials contain phlogiston" and one who does not, might consist in the fact that the first, but not the second, believes that if a flammable material were to contain phlogiston then it would contain a colorless, odorless substance that is liberated when that material is burned. As these examples show, a sentence **p** that expresses something one must believe to fully understand a given sentence **q** typically expresses something self-evident, something one believes in understanding what **p** says. A self-evident sentence is just the special case where one of the beliefs that must be involved in fully understanding it is acceptance of what that very sentence says.

¹⁸ For example: Gilbert Harman, "The Future of the A Priori," *Journal of Philosophical Research* Supplement (2003): 23-34, Michael Devitt, "Naturalism and the A Priori," *Philosophical Studies* 92 (1998): 45-65, Michael Devitt, "There Is No A Priori," in *Contemporary Debates in Epistemology*, eds. Matthias Steup and Ernest Sosa (Malden: Blackwell Publishing, 2005), 105-114.

I claim that the sentences in list (1) above are self-evident, satisfy (D1), that if someone fully understands one of those sentences (while lacking reason to believe that what it says is incoherent or is neither true nor false), then it follows that she believes what it says. It's hard to see what could show this to be wrong. It is hard to see how there could, for example, be a clear case of someone who is reluctant to accept what is said by the sentence "One plus one is two," is utterly without any idea that what it says may be incoherent or neither true nor false, but nevertheless fully understands what the sentence says.¹⁹ Until we are given reason to think that there are, or could be, cases of that sort for sentences like those in list (1), it is reasonable to hold that these sentences (and many others) do satisfy our definition of self-evidence, and that what such sentences say can therefore be known a priori.

The Contingent Self-Evident

Are there any sentences such that what they say is both self-evident and only contingently true? Consider any sentence of the following form²⁰

(4) If in the actual world there is exactly one thing that is a G, then the actual G is a G.

What such a sentence says is self-evident: failure to believe what it says would betray failure to fully understand what it says. Anyone who fully understands this

¹⁹ Alvin Goldman rhetorically asks, with respect to any simple logical truth, "Can't we conceive of psychological operations that would suffice to grasp the components and ...[composition] of ...[the proposition] but do not suffice to produce *belief* in the proposition?" (Alvin Goldman, "What Is Justified Belief?," in *Justification and Knowledge*, ed. George Pappas (Dordrecht: Reidel, 1979), 4.) Assuming that grasping the components and composition of a proposition entails fully understanding it, I myself cannot conceive of such operations for any proposition expressed by one of the sentences in my list (1), that is, I cannot imagine anything I would be willing to count as satisfying Goldman's description with respect to it. Juan Comesana, in comments on an earlier version of this paper, suggests that someone who fully understands what one of those sentences says could fail to believe it because he is irrational or lacks the motivation or capacity to believe it. But this seems to beg the question against the view that believing what such a sentence says is part of what *constitutes* understanding it: it assumes, what that view would deny, that a subject might have everything it takes to fully understand what is said by one of those sentences while lacking something needed to believe it. That view would imply instead that one who lacks the rationality or motivation or capacity to believe what one of those sentences says also lacks the rationality or motivation or capacity to understand what it says.

²⁰ Philip Kitcher, "Apriority and Necessity," *Australasian Journal of Philosophy* 58 (1980): 89-101, suggests that sentences like these express contingent a priori propositions.

conditional sentence (and such a person must have mastered some technical terms in the philosophy of language) understands that if its antecedent is true then its consequent is true. But if its antecedent is only contingently true—it is only contingently true that in the actual world there is exactly one thing that is a G (for example, let ‘G’ be “person named Carl Ginet born in 1932 who became a philosopher”)—what the whole conditional sentence says is also only contingently true. It is true in the actual world but not in any other possible world where the thing that is a G in the actual world exists but is not a G (where I exist but am not named Carl Ginet or do not become a philosopher). For in such a world the consequent is false (it is not the case that the actual G is a G there) but the antecedent is still true (it is still the case that in the actual world exactly one thing is a G).

What a sentence of form (4) says is a peculiarly uninformative sort of contingent truth. For it gives us no information that distinguishes the actual world from any other world: it is, as Gareth Evans has put it,²¹ only *superficially* contingent and not *deeply* contingent.

The superficiality of the contingency of what such a sentence says can be seen as follows. What it implies about any arbitrarily selected world *w* can be put this way:

(4a) It is true at the actual world that $\exists!xG(x) \rightarrow$ it is true at world *w* that G (the actual G).

This fails to hold for a world *w* where the actual G is not G. What (4) implies about the actual world, namely,

(4b) It is true at the actual world that $\exists!xG(x) \rightarrow$ it is true at the actual world that G(the actual G).

does hold because what (4b) says is *formally* true. And from a formal truth about the actual world we can learn nothing that distinguishes it from any other possible world. (4) is contingent because there are possible worlds where what (4) says would be false; but it is *not deeply contingent* because what (4) tells us about the actual world does not distinguish it from any other possible world.

²¹ See Gareth Evans, “Reference and Contingency,” *The Monist* 62 (1979): 161–189, from which I have derived my understanding of this distinction.

Are there any sentences such that what they say is deeply contingent and also self-evident? I can think of none that, like (4), fits (D1), the definition of self-evidence given earlier. But consider the sentence:

(5) I exist.

Because (5) contains the indexical term 'I', whose meaning is such that its referent changes depending on who utters it, definition (D1) does not apply to this sentence. But it is clear that one cannot understand what the sentence "I exist" says and not believe what it says. It certainly seems that the essence of my notion of self-evidence is present here.

If so, what definition of self-evidence for context-sensitive sentences will capture it properly? For a time I thought the following definition would do the trick:

(D3-prelim) For any declarative sentence **p** whose meaning is such that what it says may vary from one context of utterance to another, if a person *S* utters **p** in order to make an assertion and understands and believes what she thereby says, then what *S* thereby says is self-evident to *S* if and only if for any person *x*, if *x* were to utter **p** in order to make an assertion and fully understood what she would thereby say (and is "innocent", i.e. does not have reason to think that what she would thereby say is incoherent or neither true nor false), then it follows that *x* would believe what *x* would thereby say.

What this definition gives us is no doubt *a* plausible way of using 'self-evident' with respect to context-sensitive sentences. But it turns out that it is not the way I want (as Pryor has helped me to see²²).

What I want is a definition on the basis of which it will be plausible to claim that when what would be said by someone's uttering a context-sensitive sentence is self-evident to that person, then she has *a priori* (as well as non-inferential) justification for believing what would be said. But there are context-sensitive sentences that satisfy (D3-prelim) of which it would not be plausible (or at any rate I would not want) to claim that an utterer's justification for believing what she says is *a priori*.

Consider this sentence:

(6) I am uttering a sentence.

²² James Pryor, "Hyper-Reliability and Apriority," *Proceedings of the Aristotelian Society* 106 (2006): 327-344.

About this sentence Pryor says:

Given what this sentence means, it follows that whenever it's used to think a thought, that thought is true. (I count rehearsing a sentence to yourself privately as a kind of utterance.) And anyone who understands the sentence is in a position to know this. However, suppose you do utter the sentence (either privately or aloud). What then justifies you in believing that you are uttering it, or any sentence? It can't be your understanding of the sentence. That would only justify you in having beliefs about what's true *whenever the sentence is uttered*. It doesn't help you determine when that condition is fulfilled. The natural thing to say is that what justifies you in believing you are uttering the sentence is your introspective or perceptual awareness of uttering it. Hence, your justification for believing the thought you have by rehearsing (8) is a posteriori—despite the fact that you know, just by virtue of understanding (8), that whenever it's used to think a thought, that thought is true. [Underlining added.]²³

If a person utters (6) in order to make an assertion and understands what she thereby says (and is 'innocent'), it follows that she believes what she thereby says; so this sentence satisfies (D3-prelim). (But couldn't she fail to believe what it says because she fails to be aware that she has uttered it? She could, but in that case it could not be that she uttered it *in order to make an assertion*: one cannot utter something while intending by that act to make an assertion without being aware of one's act.)

The awkward thing is that, as Pryor says, at least part of what justifies the utterer in believing what she says in uttering "I am uttering a sentence" is her awareness that she is uttering it. (And there are other sentences that satisfy (D3-prelim) of which a similar thing is true, e.g., "I am uttering English," "I am uttering a sentence about what I'm currently doing".) I agree with Pryor that if one's belief that one is doing something is justified only by one's being aware that one is doing it then one's belief is not justified *a priori*: one's justification is not independent of experience in the way required for it to be *a priori*. Since (D3-prelim) makes what is said in uttering these sentences self-evident to the utterer, it is not a definition of self-evidence for context-sensitive sentences that yields the result that, whenever what is said by the utterance of a context-sensitive sentence is self-evident to its utterer, that person is justified *a priori* in believing what they've said: it is not the definition of self-evidence for such sentences that I want.

What is the definition I want? I believe it is the following:

²³ Pryor, "Hyper-Reliability," 334.

(D3) For any declarative sentence **p** whose meaning is such that what it says may vary from one context of utterance to another, if at any given time a person *S* understands and believes what she would say were she then to utter **p** in order to make an assertion (whether or not *S* then actually utters **p**), then what *S* believes is self-evident to *S* if and only if for any person *x*, if *x* understands what she would say were she to utter **p** to make an assertion (and is “innocent”) then it follows that *x* believes what she would thereby say.

This is not satisfied by sentence (6) (or by “I am uttering English” or “I am uttering a sentence about what I'm currently doing.”) One can understand what one would say by uttering that sentence at a time when one is neither uttering it nor in any other way occurrently thinking the thought that it expresses. At such a time one will not (normally) believe what one would say by uttering it, namely, that one is uttering a sentence.

Contrast sentence (5) “I exist.” There cannot be a time at which one understands that sentence and fails to believe what one would say by uttering it (or it fails to be true). Every time at which one understands it must be a time at which one believes it. This includes times at which one is neither uttering that sentence nor in any other way occurrently thinking the thought that it expresses. The same is true of some other sentences, such as “I am” or “The world contains me.”

But there are still other sentences of which it is not true, such as “I am thinking” or “I am conscious.” “I am thinking” implies that one is in a particular state such that it is possible to exist without being in that state. At a time when one exists but is not thinking one might understand “I am thinking” but not believe what it says. On the other hand, if one *is* thinking and one understands “I am thinking,” then one will of course believe that one is thinking and be justified in doing so; but one's being justified does not follow merely from the fact that one understands what one would say by saying “I am thinking;” it's required also that one be aware that one's current state is of the thinking kind. Since one's justification is in this way dependent on one's having a particular kind of experience, it should be regarded as a posteriori.

In the case of “I exist” or “I am” or “The world contains me” or any sentence such that what one says by uttering it is self-evident to one in the sense of (D-3), one's being justified in believing what one says does follow merely from one's understanding it. It is not also required that one be aware that one's condition is of some particular kind that it might not have been, for none of those sentences implies that one's condition is of any particular kind. Since one's justification is not dependent on one's being aware of being in any particular condition, one's justification should be regarded as a priori, not a posteriori.

Although what a person would say by uttering “I exist” is self-evident to the utterer (in the sense of (D-3)), and hence known a priori by them, it is also contingent. And it is deeply contingent. For in knowing at a particular time that I exist I know something that distinguishes the actual world from some other possible worlds, namely, those where I do not exist at that time. What my utterance of (5) “I exist” would say about any arbitrarily selected world w , namely,

(5a) It is true at world w that $\exists xx=me$

fails to hold for any world w where I do not exist. What my utterance of (5) would say about the actual world, namely,

(5b) It is true at the actual world that $\exists xx=me$

is true, of course, but, unlike (4b), it is not formally true. So we do not have the sort of reason we had with respect to (4b) for saying that we learn nothing from its truth that distinguishes the actual world from any other possible world. Nor do I see that we have any other sort of reason for saying so. So what I would say in uttering “I exist” is deeply contingent, as well as something I know a priori.²⁴

²⁴ What sentences (5) and (6) say when uttered is contingently true. Some simple context-sensitive sentences are, however, such that, although what they say when uttered in the right circumstances must, as with sentences (5) and (6), be believed by the utterer who fully understands what they say, what they say in those circumstances is, unlike with sentences (5) and (6), necessarily true. Consider, for example, the sentence “*That color* is maroon”, uttered by someone whose visual experience is that of a normal perceiver focussing attention on a maroon expanse and whose intention is that “*That color*” should refer directly to the *shade* of color and does not mean “the color of that object now”. That object might have been a different color now, but that shade of color could not have been other than maroon: there is no possible world where that shade is other than maroon. And an utterer whose demonstrative attention is focused on maroon must believe the necessary truth the sentence expresses if she fully understands what it says (knows what shade of color “maroon” denotes). But her justification for believing this simple necessary truth is not a priori because it depends essentially on her being visually aware of an instance of maroon. Neither, however, does what this sentence says satisfy the definiens of (D3) (or, for that matter, (D3-prelim): it is not, by (D3), self-evident to the believing utterer, despite the fact that in those circumstances she must believe it if she fully understands it. In circumstances where she is either not experiencing maroon or not uttering the sentence (or both) it would not follow from her understanding what the sentence says that she believes what it says.

A Priori Truth

Philosophers speak of a priori truth as well as a priori knowledge. We've used our notion of self-evidence to define a species of a priori knowledge. Can we use it to define a species of a priori truth? One plausible thought as to the general relation between a priori knowledge and a priori truth is that a truth is a priori just in case it is, or can be, known a priori. This might suggest that we should say that if a truth is, or could be, self-evident to someone, then it is a priori. Or, more precisely, the suggestion is that we should say the following: if a sentence **p** is such that were S to utter **p** in order to make an assertion then S would assert a truth that is self-evident to S, then the truth S would have expressed by uttering **p** is a priori.

There can be no objection to this suggestion as long as we consider sentences that satisfy definition (D1), sentences such that what they say cannot vary from one context of utterance to another, like those in list (1).

But some sentences that are such that what they say does vary from one context of utterance to another—sentences to which definition (D3), but not (D1), applies—should, I think, give us pause. In particular there seems reason to doubt that a sentence like (5) “I exist” —which would when uttered by S express a truth self-evident to S—would thereby express an a priori truth. The reason is that the deeply contingent truth S would thereby express cannot be self-evident to anyone else. There is no sentence such that were someone else to utter that sentence they would say what S says by uttering “I exist” *and* what they would say would (by D3) be self-evident to them. And there seems to be no other way that anyone else could know that truth a priori, that is, for others there is no way of knowing it of which it would be plausible to say that it is an a priori way of knowing it. And indeed there is no sentence such that were S herself to utter it *tomorrow* she would say what she says *today* by uttering “I exist” *and* what S would say by that utterance tomorrow would be self-evident to her. There seems to be no a priori way in which S could at some time other than now know the truth she would now express by uttering “I exist”. I think we should not want to classify as a priori any truth that can be known a priori only to one person at one time, the person and time the truth is about. I can know a priori that I exist (now) and each of you can know a priori that you exist (now), but none of these truths is an a priori truth.

So, although we can assert that

(T1) every a priori truth must be knowable a priori,

Carl GINET

we cannot assert the converse, that every truth knowable a priori is a priori. Might we, however, venture to claim that

(T2) every truth that is in principle knowable a priori by anyone at any time is an a priori truth,

or else that

(T3) every truth knowable a priori that is not deeply contingent is an a priori truth?

Are these perhaps equivalent? That is, is the following the case?

(*) A truth that is knowable a priori is in principle knowable a priori by anyone at any time iff it is not deeply contingent.

Consider the following sentence:

(7) There exists now at least one person.

This sentence satisfies the condition in (D3): anyone who understands what she would say by uttering this sentence must believe what she would thereby say, and so what she would say would be self-evident to her. It also follows that what she would thereby say is true and therefore something that she (then) knows a priori. But what she would say would be deeply contingent. Moreover, unlike the case with sentence (5), what she would say would be the same thing as anyone else would say by uttering sentence (7). So we have a deeply contingent truth that is knowable a priori by anyone (who exists at the time). Do we therefore have a counterexample to (*)? No. Owing to the presence of the indexical 'now,' the truth expressed by the utterance of sentence (7) at any particular time is not knowable a priori (by anyone) at any other time.

What about the following sentence?

(8) There exists now, or has existed, or will exist at least one person.

The same is true of it: owing to the presence of the indexical 'now,' the truth expressed by the utterance of (8) at any particular time is not knowable a priori (by anyone) at any other time. It is true that utterances of sentence (8) at two different times say things that are necessarily equivalent, that is, it is necessary that if either of them is true then so is the other. But the truth expressed by the one utterance is not the very same truth as the truth expressed by the other

utterance. The particular time referred to in the one utterance (by 'now') is different from the particular time referred to in the other. In a loose sense they say the same thing, i.e., the things they say are equivalent, but, since they say it about two different particular times, in a strict sense they do not say the same thing.

But consider the following sentence.

(9) There is a time such that there is at least one person at that time.

Suppose we stipulate that both occurrences of 'there is' in this sentence have no tense, so that the sentence contains no deictic reference to the time at which it is uttered (or to any other time).²⁵ Then it seems to be the case that the truth that would be expressed by someone's uttering this sentence at any time is the same as the truth that would be expressed by anyone else's uttering it at any other time, and that anyone who fully understands what she would say by uttering this sentence must believe what she would thereby say and so what she would say would be self-evident to her: the sentence expresses a truth that is knowable a priori by anyone at any time. And yet it seems also to express a deeply contingent truth: what it tells us about the actual world (or any world in which it is expressed) does distinguish that world from some other possible worlds, those where no person ever exists.

So sentence (9) seems to show that claim (*) is false and thus that (T2) and (T3) are not equivalent, that there is at least one deeply contingent truth that is knowable a priori by anyone at any time. So (T2) is stronger than (T3): (T2) entails that (9) expresses an a priori truth and (T3) does not entail this.

Suppose we think that any truth knowable a priori by anyone at any time, even one like (9) that is deeply contingent, should count as a priori. Then we can accept both (T2) and (T3), and we can accept the converse of (T2), namely:

(T2-con) Every a priori truth is one that is in principle knowable a priori by anyone at any time.

But we cannot then accept the converse of (T3), namely:

(T3-con) Every a priori truth is one knowable a priori that is not deeply contingent.

²⁵ Thanks to Sydney Shoemaker for suggesting that I consider this sentence.

If we accept (T2) then we are committed to accepting that (9) expresses an a priori truth, but that contradicts (T3-con) since the truth that (9) expresses is deeply contingent. If, on the other hand, we think that the truth (9) expresses should not count as a priori, then we must reject (T2) but we can accept (T3) and both (T2-con) and (T3-con).

Which package should we accept? I have no clear intuition that one or the other is correct. I waver. Is it really plausible to count a deeply contingent truth as a priori? Well, why not, if it is in principle knowable a priori by anyone at any time? In the absence of a compelling intuition one way or the other, perhaps what is needed is a decision. But I have no clear preference, no clear sense that one way rather than the other would, all things considered, be the better way to go.²⁶ Rather than rush to an arbitrary decision, I think that I will just leave it open for now and solicit advice.²⁷

²⁶ I do think that either of these alternatives is preferable to a third alternative of making the application of a priority to truths relative to the subject and/or time the truth is about, so that, e.g., sentence (5) expresses a truth that is an a priori truth *for* the subject and time it is about, but not an a priori truth for anyone else or at any other time (a suggestion broached by Max Deutsch in personal communication). That is, I don't like this as a way of explicating how the term 'a priori' applies to truths as such; if, however the suggestion is simply to abandon applying the term to truths as such and to confine ourselves to talking only about who can, and who cannot, know a truth a priori, then it might be worth considering.

²⁷ I am grateful for helpful comments I received from Sydney Shoemaker and Matti Eklund and from discussion participants when I presented earlier versions of this paper to the Epistemology and Metaphysics Discussion Group at the University of Toronto, to the Philosophy Department at Davidson College, to the Conference on the Epistemology of Keith Lehrer at the PUCC, Porto Alegre, Brazil, and to the Philosophy Departments at the University of Sussex, Hong Kong University, The Chinese University of Hong Kong, Hong Kong Polytechnic University, and Lingnan University. An earlier version of this paper appeared in the Brazilian journal *Veritas* 54, 2 (2009).

NO TIME TRAVEL FOR PRESENTISTS

Steven D. HALES

ABSTRACT: In the present paper, I offer a new argument to show that presentism about time is incompatible with time travel. Time travel requires leaving the present, which, under presentism, contains all of reality. Therefore to leave the present moment is to leave reality entirely; i.e. to go out of existence. Presentist “time travel” is therefore best seen as a form of suicide, not as a mode of transportation. Eternalists about time do not face the same difficulty, and time travel is compossible with eternalism.

KEYWORDS: presentism, time travel, time, metaphysics

Presentism is incompatible with time travel. Thus far, the literature on their compatibility has focused on one argument, the so-called Nowhere Argument. According to The Nowhere Argument, if the present is all that is real, then there is nowhere for a would-be time traveler to go, and so time travel is not possible for presentists. This argument has been roundly dismissed by recent commentators. The present paper presents a new reason to reject the compatibility of presentism and time travel, namely, the Suicide Machine Argument. It will be shown that the moves that presentists make to get around the Nowhere Argument are not successful to fend off the Suicide Machine Argument.

Presentism

Here are some recent statements of presentism:

- Only the present exists.¹
- Only currently existing entities exist, and... the only properties and relations those entities instantiate are those they currently instantiate.²

¹ Craig Bourne, *A Future for Presentism* (Oxford: Oxford University Press, 2006), 13. Bourne’s complete theory of presentism is complicated: he defends an “ersatz presentism” according to which times are abstract objects; they are ordered pairs of present-tensed propositions that do not contain *past* or *future* operators, and a date, ordered by a relation isomorphic to “earlier than”. Abstract objects aren’t in time, therefore can’t be ordered by the true “earlier than” relation. The present is the only concrete instantiation of a time. Bourne’s theory is a shadow eternalism behind a concrete presentism.

- It is always the case that for every x , x is present.³
- Presentism... is the doctrine that all reality is confined to the present—that past and future things simply do not exist, and that all quantified statements that seem to carry commitment to past or future things are either false or susceptible of paraphrase into statements that avoid the implication.⁴

Bourne, Sider, and Zimmerman employ tenseless quantification, and most presentations of presentism tend to do so. The reason is easy to see: “everything exists [present tensed] exists now” is trivial, since it means no more than whatever exists now, exists now. Crisp tries to work out a nontrivial definition of presentism that involves tensed existence, but that complication is not important for the present discussion. Presentists agree that there may be things that do not exist in time, like abstract objects or God, but the root presentist idea is that everything that exists in time is *simultaneous*. You can’t have (tenselessly) existing things at different places in time. Everything that exists, exists at once.

Presentists are committed to a purely objective present; there is a non-relative fact about the stuff of the world at one time that they are present/happening/real, and other events and objects merely were or will be. Eternalists, on the other hand, are committed to a merely subjective present; the fact that such-and-such date (399 BCE, 2010, 2011, whatever) is present can only be true relative to something or some event, and is not true in an objective or nonrelative way. That is, ‘present’ like ‘here,’ ‘this,’ or for modal realists ‘actual,’ is indexical. There isn’t a unique present.

Truthmakers and the Nowhere Argument

The reason that some have doubted the compatibility of presentism and time travel is that if the present is all that is real, then Apatosaurus is just as nonexistent as King Kong. One cannot journey to the nonexistent, and so if presentism is true, then one cannot travel to nonexistent points in time, namely, any time other than the present. The prospects of traveling in time to warn Socrates off the hemlock are no more viable than traveling to Hogwarts to warn Harry Potter about Lord

² Theodore Sider, “Traveling in A- and B- Time,” *The Monist* 88, 3 (2005): 329.

³ Thomas M. Crisp, “Presentism,” in *The Oxford Handbook of Metaphysics*, eds. Michael J. Loux and Dean W. Zimmerman (Oxford: Oxford University Press, 2003), 215, cf. Thomas M. Crisp, “On Presentism and Triviality,” in *Oxford Studies in Metaphysics, Volume 1*, ed. Dean W. Zimmerman (Oxford: Oxford University Press, 2004), 16.

⁴ Dean W. Zimmerman, “The A-Theory of Time, the B-Theory of Time, and ‘Taking Tense Seriously,’” *Dialectica* 59, 4 (2005): 402.

Voldemort. Simon Keller and Michael Nelson call this the Nowhere Argument; for a presentist, time travel is a nonstarter because there is nowhere to go.⁵

The Nowhere Argument is essentially a time-travel variant of the broader truthmaker argument against presentism. According to the truthmaker argument, if the present alone is real, then no statement about the past or future can have a truth value. This is because there is nothing about reality that would make such a statement either true or false. True propositions must have truthmakers, and there is no past or future reality to make a statement about the past or future true, *mutatis mutandis* for false statements. Yet of course it is perfectly true to say that “Pete Sampras won the Wimbledon’s men’s final in 1999 against Andre Agassi” and “Roger Federer will win several tournaments over the next few years.” Since presentism cannot allow such sentences to have any truth value at all, presentism is false.

The truthmaker argument has been recently defended by Michael Dummett in his Dewey Lectures.⁶ The canonical reply of presentists is to respond that the truthmakers for past and future facts are located in the present, and are expressed by appropriately tensed language.⁷ As John Bigelow states it, the truthmaker principle may insist that truth supervenes on being, but the presentist can aver that there are present intrinsic properties of the world upon which all past and future tensed facts supervene. Moreover, sentences that quantify over times can be translated without remainder into tensed sentences with present truthmakers. There’s some finessing that presentists need to do about singular statements involving objects that no longer exist, such as “Socrates was snub-nosed,” since there is nothing in the presentist’s universe that ‘Socrates’ refers to. And it is hard to see how sentences with non-referring subject terms can be true. In this case, presentists tend to take a page out of the Quinean handbook (see the section on ‘Pegasizing’) and start talking about individual essences, or haecceities.⁸

⁵ Simon Keller and Michael Nelson, “Presentists Should Believe in Time-Travel,” *Australasian Journal of Philosophy* 79, 3 (2001): 333–45.

⁶ Michael Dummett, *Truth and the Past* (New York: Columbia University Press, 2004), 74.

⁷ This response can be found in John Bigelow, “Presentism and Properties,” in *Philosophical Perspectives 10: Metaphysics*, ed. James E. Tomberlin (Oxford: Blackwell Publishing, 1996), 35–52; Michael C. Rea, “Four-Dimensionalism,” in *The Oxford Handbook of Metaphysics*, eds. Michael J. Loux and Dean W. Zimmerman (Oxford: Oxford University Press, 2003), sec 4; Bourne, *A Future*, 56–60.

⁸ See Ned Markosian, “A Defense of Presentism,” in *The Oxford Handbook of Metaphysics*, eds. Michael J. Loux and Dean W. Zimmerman (Oxford: Oxford University Press, 2003), 47–82; Rea, “Four-Dimensionalism;” Gary Rosenkrantz, “An Epistemic Argument for Enduring Human

Keller, Nelson, and Ted Sider all reject the Nowhere Argument on similar grounds. In the case of time travel, a presentist will translate sentences with ostensible commitments to times into sentences with tensed expressions. Here's Sider:

Instead of claiming that *there exists* a dinosaur-viewing by me, located two hundred million years before the present time, the presentist can say "it was the case two hundred million years ago that I am viewing a dinosaur." Instead of ascribing a two-place causal relation to the events *my entry into the time machine* and *my viewing a dinosaur*, she can use a two-place tense operator 'because Φ , it was the case n units of time ago that Ψ ' in the following tensed claim: "because I entered a time machine, it was the case two hundred million years ago that I am viewing a dinosaur."⁹

According to the presentist, while other times besides the present moment aren't real, there are past, present, and future-tensed facts, and tense is not reducible to times. Let's assume that the appeal to tensed facts is a perfectly adequate response to the truthmaker objection to presentism. Let's even assume that, if the Nowhere Argument is just a time-travel variant of the truthmaker objection that the solution of tensed facts does the job in defeating it. The problem is that there is still a remaining conundrum for the would-be presentist time traveler, one that cannot be resolved though the clever manipulation of tensed sentences.

The Suicide Machine

Let us consider what happens when Dr. Who steps into a blue police-box shaped time machine, twiddles some knobs, and disappears. Presentists and eternalists agree about the following facts.

First, the time machine is the cause of an effect at a discontinuous point in external time. That is, when viewed from the point of view of time external to the time traveler, causation is discontinuous. Dr. Who's flipping a coin now, right before he pushes the buttons in the time machine, causes it to come up heads 1000 years in the future without the coin's remaining airborne and spinning for a millennium. In the time traveler's personal time, with its forward-ordered

Persons," *Philosophy and Phenomenological Research* 57, 1 (2005): 209-24. And, of course, Willard V. O. Quine, "On What There Is," in his *From a Logical Point of View* (Cambridge: Harvard University Press, 1953), 1-19.

⁹ Sider, "Traveling," 232.

psychological states, the coin flips for only a couple of seconds before coming up heads. Time travel produces disjoint causation, a sort of action-at-a-(temporal)-distance. If the time travel is into the past, then it is a form of backwards causation.¹⁰

Second, time machines produce temporally disconnected objects. If Dr. Who punches the buttons for 399 BCE, then Dr. Who is no longer in 2010, but he was in 399 BCE, with all the physical and mental characteristics he will have in 2010. Objects are temporally scattered: either they have temporal parts that are separated by swaths of time (for perdurantists) or they wholly exist at different moments in time without passing through the moments between (for endurantists).

Now we come to disagreement. For presentists, getting into a time machine is suicide—the occupant goes out of existence. Recall that presentists are committed to a purely objective present; the events and objects at this objective present alone are real, even if other things have been or will be real. After entering the time machine, Dr. Who no longer exists in the objective present, and therefore he is no longer in reality. Which is just to say that Dr. Who ought to view the time machine with considerable trepidation—after all, it means his annihilation. If Dr. Who ‘travels’ five years into the future, then he goes out of existence now, is nonexistent for five years, but will come back into existence five years from now. In this case, “time traveling into the future” is best described as death and resurrection. It is a theologically neutral, but metaphysically loaded, resurrection of the body. Permanent “time traveling into the past,” on the other hand, has a less rosy outcome. It is merely death.

Consider H. G. Wells’s time traveler. Wells writes, “he may even now —if I may use the phrase—be wandering on some plesiosaurus-haunted Oolitic coral reef, or beside the lonely saline lakes of the Triassic Age.” For a presentist, that is *not* an apt description of the situation. The time machine may have caused it to be the case that the traveler *was* wandering on some plesiosaurus-haunted Oolitic coral reef, or beside the lonely saline lakes of the Triassic Age, but of course he isn’t doing so now. Furthermore, no part of reality contains the time traveler doing any such wandering. Whatever backward causation the time machine may have effected, the traveler is no longer in the objective present and therefore no longer exists. There is nothing farther requisite to make him a perfect non-entity.

¹⁰ Although not all presentists think that backwards causation, and hence backwards time travel, is possible. See Bourne, *A Future*, 134.

Not only does presentist “time travel” merely require the would-be traveler to go out of existence *in nihilum*, but it also requires that objects come into existence *ex nihilo*. When Dr. Who steps into the time machine in 2010 and pushes the buttons for 399 BCE, the time machine causes it to be the case that in 399 BCE, Dr. Who came into existence. But he literally came into existence out of nothing. When 399 BCE was the present, it was true that “in 2010 Dr. Who will enter a time machine, and this machine will cause him to come into existence now” (assuming a closed future with respect to Dr. Who’s adventures). However, at one second in 399 BCE, reality does not contain Dr. Who in any fashion and the next second it does contain him, whole, fully grown, with memories of his future life. The best description is that he came into being out of nothingness.

Eternalists balk at the notion that time machines are really suicide machines. They insist that any sort of successful travel, spatial or temporal, involves the traveler existing at departure and safely arriving, intact and still in reality, at the arrival. Presentist time machines won’t take you to the past, they only provide travel to the Great Beyond. An eternalist time machine is far less threatening. Eternalists deny that there is an objective present; the now, they say, is purely subjective. Socrates is every bit as real and existing as the reader of this article, and he is in the present—his own present, one that also becomes the present of Dr. Who when he sets the controls to 399 BCE.

For eternalists, Wells’s description makes perfect sense. The time traveler is (tenselessly) wandering on some plesiosaurus-haunted Oolitic coral reef, not in our subjective now, of course, but in his own subjective present. When Wells’s time traveler, or Dr. Who, bops around time, they do not go out of existence; there is no death and resurrection, no changing the past while putting oneself in the grave (albeit an empty-casket burial). The time traveler remains in reality, fully existing at every stop along the way. So there is no coming into existence *ex nihilo* either—when Dr. Who travels to 399 BCE, he is never nonexistent, and therefore does not come into being out of nothing upon arrival. He may come out of the future, but he does not come out of the void.

Presentists argue that for Dr. Who a certain collection of properly tensed facts are true of him (and whatever else is in his personal time) that are not in keeping with a set of tensed facts ordered along external time, and that this counts as being a time traveler. To be sure, Dr. Who is a strange sort of temporally disjoint object when viewed from external time (and everything else is oddly disjoint when viewed from his personal time), but being this kind of object is insufficient to count as traveling in time. Either presentists must identify the objective present with the present of external time or the present of Dr. Who’s

personal time. Suppose they identify the present of external time as the objective present (an assumption made for presentation purposes above). In this case, by the Suicide Machine argument, Dr. Who leaves the objective present and thus goes out of existence. Suppose instead presentists identify Dr. Who's personal time as the objective present. In this scenario, after he gets into the time machine everything else in the world is separated from the objective present, and hence the universe minus Dr. Who goes out of existence.¹¹ If they count both as the present, that is either to admit that the present is subjective—i.e. eternalism is true—or it is to endorse the notion of multiple objective temporal dimensions. Obviously the first horn is capitulation to eternalism. Let us consider the second horn.

Suppose that there is more than one temporal dimension, so that 399 BCE is the present, and also 2010 is the present, but on a different temporal metric. A time traveler might move from one timeline to a point on an orthogonal timeline in a way that it counted as traveling in time. Both points would (presumably) be the present at which all of reality resides, although they would not be identical with each other. There was a brief flurry of discussion of two-dimensional time in the 1970's¹² and since then the little that has been written has been rather tepid about multidimensional time.¹³

It is difficult to know what to say about this possibility, mostly because there isn't an explicitly presentist defense of it to examine. There is the following dilemma for multidimensional presentists to overcome: either (1) any such defense would have to explain how every existing thing is at point t_a on one timeline, and yet every existing thing is also at point t_1 on a different timeline without contradiction, and why the proper description is not that reality is wholly at the ordered pair point (t_a, t_1) , or (2) if reality is at an ordered pair point, then there is

¹¹ An option which may lead to the solipsism of Robert A. Heinlein's time travel story "All You Zombies."

¹² For: Jack W. Meiland, "A Two-Dimensional Passage Model of Time for Time Travel," *Philosophical Studies* 26 (1974): 153-73; T.E. Wilkerson, "Time and Time Again," *Philosophy* 48, 184 (1973): 173-77; T.E. Wilkerson, "More Time and Time Again," *Philosophy* 54, 207 (1979): 110-12. Against: Ronald E. Nusenoff, "Two-Dimensional Time," *Philosophical Studies* 29 (1976): 337-41; Ronald E. Nusenoff, "Spatialized Time Again," *Philosophy* 52, 199 (1977), 100-01. Skeptical but officially neutral: David Lewis, "The Paradoxes of Time Travel," *American Philosophical Quarterly* 13 (1976): 145-52.

¹³ Graham Oppy straightforwardly opposes it, and Murray MacBeath is skeptical but neutral. See Graham Oppy, "Can We Describe Possible Circumstances in Which We Would Have Most Reason to Believe that Time is Two-Dimensional?," *Ratio* 17, 1 (2004): 68-83 and Murray MacBeath, "Time's Square," in *The Philosophy of Time*, eds. Robin LePoidevin and Murray MacBeath (Oxford: Oxford University Press, 1993), 183-202.

the Lewis/Sider objection that no kind of travel in a two-dimensional time plane constitutes travel into one's own past. If $time_1$ and $time_2$ are imagined as axes on a Cartesian coordinate plane, and temporal movement is seen as advancing along these axes away from the origin, then backwards time travel along one of the axes will still be forward time travel along the other. In which case there is no backwards time travel to the same point in the plane.¹⁴

Another potential complaint from the presentist is that we all travel in time—we are all traveling into the future at the well-publicized rate of one second per second. If the Suicide Machine argument is right, then our own ordinary time travel into the future is just as impossible as anything out of H.G. Wells. Since it obviously possible, simple *modus tollens* shows that something has gone wrong with the Suicide Machine argument.

Here the right response is to turn the tollens into a ponens. Presentists should not take seriously talk of ordinary persistence being a sort of traveling in time. Ordinarily, time travel is understood as the removal of the traveler from the usual flow of time and jumped earlier or later in the stream—thus the Lewisian distinction between personal and external time. To count persistence along the moment-to-moment flow of time itself as time travel is to erase personal vs. external time and so to give a quite idiosyncratic interpretation of traveling through time. Moreover, there's a considerable literature on how it is possible for presentists to permit any sort of persistence (with David Lewis famously denying that they can), which is a component of an even broader literature on how it is possible for presentists to offer an adequate theory of change and solve the problem of temporary intrinsics. While it is beyond the scope of this paper to examine these issues, however presentists ultimately construe change, additional argument will be needed to show that the moving present is in some important sense a sort of time travel. Presentists need caution here, since, as argued above, for them time travel is impossible!

In the end it seems that, barring a successful presentist theory of multidimensional time, there is no such thing as time travel under presentism. Whatever the merits of the appeal to tensed facts as a general strategy against the truthmaker/ Nowhere Argument, it won't help the time travel case. Reason: for presentists, "time travel" is really suicide. Whatever odd causal links a time machine might produce, there is no such thing as traveling to other times. Unless, of course, you're an eternalist.¹⁵

¹⁴ See Sider, "Traveling," 334.

¹⁵ Thanks to Dean Zimmerman for a very helpful discussion about an ancestor of this paper.

GETTING GETTIER'D ON TESTIMONY

Lauren J. LEYDON-HARDY

ABSTRACT: There are noncontroversial ways in which our words are context dependent. Gradable adjectives like 'flat' or 'bald', for example. A more controversial proposition is that nouns can be context dependent in a reasonably similar way. If this is true, then it looks like we can develop a positive account of semantic content as sensitive to context. This might be worrying for the epistemology of testimony. That is, how can we garner knowledge from testimony if it's the case that, though our syntactic utterances are *identical*, the semantic content of them may fail to be uniform? What if we mean different things by the same words? I argue that these kinds of semantic divergences provide the groundwork for a new kind of Gettier case. That is, given the likelihood of divergent semantic content, we can see a way to scenarios in which, despite that the semantic content is uniform, we might get justified true beliefs that nevertheless fail as knowledge. This, because it just as likely could have been the case that relevant contexts were dissimilar, and thus relevant semantic content would have been divergent. Lastly, where the phenomenon does occur, we never would have known the difference.

KEYWORDS: Syntax, Semantics, Context Sensitivity, Epistemology, Testimony, Gettier

In this article I will argue that we routinely fail to transmit knowledge by testimony for Gettier-type reasons.¹ I argue for the plausibility of a broadly construed context sensitivity for semantics. The claim is that it is possible for speakers to intend, by the very same words, to express divergent propositions. Moreover, I claim that assuming syntactic uniformity, those semantic divergences easily fail to become obvious. Where we mean different things by the same words, we generally don't notice that this is the case, without doing some heavy clarificatory lifting. Thus, it follows from a very general kind of context sensitivity for semantic content, and the divergences that result from it, that these failings in communication are, perhaps even standardly, opaque. Where they occur we fail to impart knowledge by testimony. That is, if I mean to testify that *p*, and you

¹ Edmund Gettier, "Is Justified True Belief Knowledge?," in *Epistemology: Contemporary Readings*, ed. Michael Huemer (London and New York: Routledge, 2002), 444-446.

understand me to be testifying that q , despite that the syntactic form of the testimony is uniform across our interlocution, then you have learned neither p nor q from me. In addition, and given the plausibility of the opaque cases, I argue that we ought to think that even when speakers successfully share semantic content, they may well not have, and never noticed. Therefore when semantic content is uniform across speakers, it is in this sense fortuitous, the accident of its uniformity itself being opaque. It follows that when we appear to have successfully transferred knowledge by testimony, we have in fact been Gettier'd on those transfers, since it may well have been the case that we had meant different things by the same words, without ever having noticed.

In some cases it is uncontroversial that we mean different things by our words. Take for example indexicals (like 'I,' 'he,' 'they,' etc.) or gradable adjectives (like 'flat,' 'bald,' 'far,' and so on). These are examples of our contexts informing the meaning of our words—"the table is flat" can be true in one context, and not so in another. When context becomes a relevant factor in determining speaker meaning across interlocutors, it generally seems that we plainly share that context, such that its significance is obvious. When context is shared in an obvious way the efficaciousness with which we communicate arises not just from sharing a language, but also from a mutual base of reference points driven by the relevant context and employed with all candidness by speakers therein. Alternatively, when speakers fail to share a context it seems clear to all persons at hand that this is the case, such that speakers correct to account for that ambiguity.

The problematic cases begin with unapparent context sensitive meaning. I'll show how this is possible in (at least) two ways.

I. Two Base Cases

Can nouns be context sensitive similarly to indexicals and gradable adjectives? It would be interesting if they could be. For if context sensitivity is so broadly relevant to assertions then utterances are potentially *drastically* dissimilar. So much so that instances of syntactically identical utterances across interlocutors might still be semantically divergent. Though we might utter exactly the same sentence, we may take ourselves to be saying something different from our conversational counterpart. The schism in meaning that this possibility belies leverages a serious challenge within the epistemology of testimony.

One way this might come about is in instances where interlocutors have 'unique contexts.' By this I mean that their individual contexts differ sufficiently insofar as the contexts in fact inform their semantics idiosyncratically. Arguably, this can happen a number of ways. I'll assume a fairly straightforward account of

differing backgrounds to account for the idiosyncratic semantics. In the 'unique contexts' case I'll show how these dissimilar contexts may result in false beliefs from testimony.

Following the 'unique contexts' case, I will show how one might get Gettier'd on testimony. Here the standard Gettierizing 'accident-operator' is built to be the interlocutors' relevant contexts. That is, it just as likely could have been the case that, given their contexts, their semantic content (and what they took themselves to be saying) didn't match up. However, in some cases despite divergent contexts, interlocutors can still manage to communicate effectively. In any standard epistemological circumstance if a justified true belief arises by some accident, we find our intuition is that it just isn't knowledge. In standard Gettier cases this just-as-likely factor is traditionally external to the subject. Here we find that a new kind of Gettier case presents itself where the just-as-likely factor is this: under utterly normal circumstances two interlocutors can effectively mean different things by their identical utterances. Moreover, they might not know it because it just isn't obvious (the utterances are after all syntactically identical) given that the divergence across context isn't obvious. If that's true then the worry that we're (maybe even *often*) talking around one another becomes salient. If the worry is salient then it becomes an epistemological defeater for garnering knowledge from testimony. Therefore, we routinely fail to transfer knowledge via testimony for Gettier-type reasons.

1.1 Unique Context & False belief

Meet Bronwyn and Faye. Bronwyn grew up in Hudson Bay, Saskatchewan – a highly forested area with an estimated average of 600 trees per acre. The farm on which she was raised is just outside of town, and in fact entirely isolated by the surrounding forests, which are made up mostly of fir trees and paper birch, made thicker still by the tall growing bushes of Saskatoon berries. As a young adult Bronwyn moves away to live in New York City, where she studies as an undergraduate at NYU. On her dorm floor lives Faye, who also moved to New York, but from Texas. Faye grew up all her life in the southern states, and, prior to leaving home at eighteen, had never so much as seen the kind of greenery that surrounded Bronwyn completely, before coming to New York. The two make fast friends over their first year at NYU and in the summer Bronwyn invites Faye back to her farm to meet her family. A few days into their Canadian get-away, the following conversation takes place:

Faye: What a beautiful place. It must have been a wonderful to grow up here.

Bronwyn: It was. You know, I built a tree house out here when I was a little girl.

Faye: Oh, yeah? Where about?

Bronwyn: Just north of the house, at the edge of the forest.

Faye: Hmm. I was just walking in that forest this morning and I didn't see your tree house. It must be gone now.

Bronwyn: It's not gone. In fact, I just came from there. Are you sure you were in the north forest?

Faye: Yes, I was in the north forest. I guess there is a tree house at the edge of the forest. I simply must have missed it.

Bronwyn's experiential evidence informs the semantic content—the meaning—that maps to the syntactic structure of 'forest.' Assume that when Bronwyn utters 'forest' she means 'a cluster of trees with a minimal density of 600 trees/acre.' Faye, however, having grown up in Texas all her life, and only just moving to New York, will have a drastically different experiential background for the content formation of 'forest.' When Faye utters 'forest' she means 'a cluster of trees with a minimal density of 100 trees/acre.'

Assume the 'north forest' is the kind of forest that increases in density as you venture further into it, as forests tend to. Where the first trees appear its density is 100 trees/acre, which we know is sufficient to satisfy Faye's semantic content for 'forest.' However, for Bronwyn the forest proper will not count as having started until the density of the trees reaches 600/acre, in fact several meters away from Faye's 'edge of the forest.' Does Faye know that there is a tree house at the edge of a, or the, forest?

Part of the difficulty with saying that Faye knows that there is a tree house at the edge of 'the' forest begins with her evidence to the contrary. What makes that evidence palpable, moreover, is its predication on her contextually unique semantic content for the syntactic utterance of 'forest.' Considering the less stringently dense Faye-forests, the proposition that Faye took Bronwyn to express is false.² For Faye

² Jason Bridges worries that these kinds of miscommunications boil down to someone simply being *wrong*. For example an interlocutor might speak truly in conversation, prompting another interlocutor to make some justified knowledge assertion that's yet easily answerable as false. It might be case that the kinds of miscommunications we're worrying about here are simply the result of a misapplication of terms, rather than a more troubling difference in legitimate semantic content, but it's difficult to see who is wrong in the case of Faye and Bronwyn, partly

to go back and double-check we remain certain that she would find no tree house at the edge of the 'forest,' since where she will look for the edge of the forest will in actual fact be a different place from the place to which Bronwyn is referring. Moreover, everything that counts as a forest for Bronwyn will in a sense count as a forest for Faye. By virtue of Faye's idiolectically weaker standards for forest hood, all Bronwyn forests slide into that Faye-inclusive group since they're more than sufficiently dense. Contra positively, not all Faye-Forests will qualify for forest-hood on Bronwyn's standards. The question of whether or not the tree house is at the edge of *a* forest is only true in virtue of Bronwyn-forests. Thus, Faye doesn't know that the tree house is at the edge of the, or a, forest. It is also *actually* false that the tree house is at the edge of the, and a, forest on Faye's standards, but actually true that the tree house is at the edge of the, and a, forest on Bronwyn standards.³

I.II Unique Context & Accidentally True Belief

Case two is a standard Gettier case.

Lucas: Faye tells me there's a forest behind your family home.

Bronwyn: Yeah, and there's a tree house at the edge of that forest.

Faye: Oh yes, there's a tree house at the edge of that forest.

Lucas: Oh, did you see it?

for the ship of Theseus-type reasons. When does it start and stop being a forest? Since this isn't clear, I'm comfortable maintaining that there are at least these cases where divergent semantic content is unproblematically a phenomenon that can result in interesting miscommunications, without either interlocutor being flatly wrong. See Jason Bridges, "Wittgenstein vs. Contextualism," in *Wittgenstein's Philosophical Investigations: A Critical Guide*, ed. Arif Ahmed (Cambridge: Cambridge University Press, 2010), 109-128.

³ It's conceivable that Bronwyn include something like, "Wait—you know that I mean by 'edge of the forest' the point at which the trees are dense enough to count as a 'forest,' right?" This kind of interlocation might not be enough to change Faye's standards for forest-hood, but would almost certainly help to clarify the object of reference in the conversation. Alternatively, interlocutors might be motivated by more pressing or immediately relevant circumstances to exercise exhaustive clarity. For example, if Bronwyn and Faye were signing a contract with respect to forests, the meaning of the locution would be more explicitly defined. Of course naturally conversations don't translate to the clarificatory demands of contracts.

Faye: No, I've seen the forest but not the tree house at its edge. But there is a tree house at the edge of that forest. Bronwyn told me so.

We know from the first case that the semantic content assigned to 'forest' by Bronwyn and Faye differs by 500 trees/acre. It follows that the edge of the forest for Faye will be several meters out from the inner most forest, where the trees will be dense enough to qualify as a Bronwyn-forest. However here, unbeknownst to either Faye or Bronwyn, on the opposite side of the Faye-forest there is another tree house. So, on the south facing edge of the Bronwyn-forest is the tree house of which Bronwyn speaks, and on the north facing edge of the Faye-forest there happens also to be a tree house. Thus, when Bronwyn says (using Bronwyn standards for forest-hood) that there is a tree house at the edge of the forest behind her family home, she speaks truly. Thus, the information that Bronwyn's testimony imparts to Lucas is true. Moreover, when Faye concurs with Bronwyn that there is a tree house at the edge of the forest behind her family home, she too speaks truly, even on Faye standards for forest-hood. But, that Faye's assertion is true is merely so in virtue of facts unbeknownst to herself. Neither she nor Bronwyn are aware of the second tree house on the edge of the Faye-forest. Faye's belief that her assertion is true is justified because of Bronwyn's testimony. However the tree house that actually makes Faye's assertion true is alien to Bronwyn. Thus, the truth of Faye's assertion cannot appeal to Bronwyn's testimony for its justification.⁴

Does Faye know that a tree house is at the edge of *the* forest? Yes and no. Because every Bronwyn-forest is trivially a forest for Faye, the tree house to which Bronwyn refers is on the edge of some forest. But, it's not the same tree house that makes Faye's assertion true. Without radical concessions, if Faye knew which tree house Bronwyn was referring to, she wouldn't assent to the tree house being at the edge of a forest. It might seem better said that Faye knows that a tree house is at the edge of *a* forest, but this too is peculiar. Who *doesn't* know that *somewhere* there is a tree house at the edge of a forest? Or, if Faye knows that a tree house is at the edge of a forest because the actual tree house that makes her assertion true is a separate tree house from the one to which Bronwyn refers, then exactly what kind of epistemic connection can be drawn between the second tree house on the edge

⁴ I am here leaning on a traditional account of epistemological testimony that trades on the transmission principle. That is, one cannot impart knowledge without first having knowledge (justified true belief) of the propositional content. Jennifer Lackey, in *Learning from Words: Testimony as a Source of Knowledge* (Oxford: Oxford University Press, 2008), proposes several challenging counter examples to this thesis. However, given space constraints it will suffice to assume the limitations of the traditional account here.

of the Faye-forest and Faye's knowledge assertion? The justification of her belief leans on Bronwyn's testimony. Its truth, however, is fortuitous. Thus we will say that Faye's been Gettier'd.

II. Accidentally Shared Contexts & True Beliefs

The transmission principle for testimony tells us that the truth of what is believed by the hearer must match up with the truth of what's been asserted by the utterer. I argue that it's *easy* for it to match fortuitously. This is true because of the context dependence of an utterer's semantic content. Our contexts are easily divergent for one reason or another. It follows from this that our semantic content is easily divergent, in virtue of our contexts. Where our semantic content is divergent, the truth of our utterances can easily match fortuitously.

What exactly gets lost in interlocution when our semantic content is relevantly divergent? I have tried in the base cases to illustrate some examples. Specifically, I have tried to show how the referent itself can go astray. In what remains I hope to show how the referent can remain whilst we nevertheless fail to preserve knowledge. I turn now to cases of accidentally shared contexts. My hypothesis is that although semantic uniformity can be provided by accidentally shared contexts, and thus may preserve justified true beliefs, the case of accidentally shared contexts will nevertheless fail to provide for testimonial knowledge due a hybrid Gettier-type concern about belief and meaning.

II.I The New Gettier: An Argument

Suppose that two interlocutors share a relevant context, and thus share semantic content. Suppose further that the context is only accidentally shared. Because the context is only accidentally shared, the semantic content is only accidentally shared. Given that the context is *accidentally* shared (it just as likely could have been the case that their contexts, and thus their semantic content, were divergent) *any* resultant belief is disqualified from knowledge on Gettier-principles. By 'any' I mean that the resultant belief need not appeal to its truth or falsity to determine its epistemic eligibility. It doesn't even get that far. The problem of accidentally shared contexts is logically prior to the truth or falsity of the resultant belief. The problem is that there doesn't seem to be any *prima facie* discernable evidence to distinguish in real cases between unique contexts resulting in accidentally true beliefs, and accidentally shared contexts resulting in true beliefs. Yet one seems more worrying than the other. In the latter true or false beliefs don't matter—any belief from an accidentally shared context cannot amount to knowledge. But if it's not obviously

discernable whether a case is accidentally or verifiably contextually (and thus semantically) uniform, how can knowledge from testimony ever be preserved?

Mutually shared contexts make knowledge more likely, but they cannot in principle guarantee that we garner knowledge. This is true even when the object of the knowledge ascription obtains, semantically speaking, because:

The New Gettier (TNG): We routinely fail to impart knowledge through testimony for Gettier-type reasons.

If backgrounds might be shared only accidentally then knowledge from testimony requires more than semantically identical references across interlocutors with shared backgrounds. Where backgrounds are shared accidentally interlocutors will preserve uniform semantic content. There are lots of ways for interlocutors to only accidentally share context. Perhaps the situation is such that the odds of two speakers sharing some relevant background are sufficiently low, or maybe for some reason there's simply no way of verifying that there is a relevantly shared context at all. In cases where interlocutors either can't check, or, if they could, probably wouldn't share a context (though might still), we would say they accidentally share that context. To accidentally share a context is just to say that, a) it just as easily could have been otherwise and, b) we more than likely wouldn't have noticed. They are thus Gettier'd on the transfer of knowledge through testimony.

Here is the argument for TNG:

- 1) Contexts are often only accidentally shared.
- 2) When context is only accidentally shared, the truth of the resultant beliefs only accidentally matches the truth of the utterance, since
- 3) For knowledge to transfer, the truth of what is uttered cannot only accidentally match to the truth of what is believed.

Therefore, c) We routinely fail to impart knowledge through testimony for Gettier-type reasons.

II.II The Linguistic Defeater

Suppose that Faye is wandering about town on her own when she meets Joanne. This morning Bronwyn has described to Faye the tree house that she built at the edge of the forest as a child. However, Bronwyn is unsure that the tree house remains. Bronwyn seems nostalgic for her tree house of old, and so Faye goes looking for it. In actual fact the tree house that Bronwyn built is gone. There is however a new tree house, located at the edge of the forest where the trees are dense enough to satisfy Faye's conditions for forest-hood. Joanne, unbeknownst to Faye, is new in town. By pure chance, she's from Texas too.

Faye: Excuse me? My girlfriend tells me that at the edge of the forest there is a tree house that she built when she was younger.

Joanne: I did see a tree house at the edge of the forest, just the other day.

Faye: So there *is* a tree house!

Joanne: Oh yes, there's a tree house at the edge of the forest.

It looks like there's a defeater for Faye's belief. Namely, that lots of people around could just as well have meant something totally different by 'forest.' Joanne happens to share Faye's standards for forest-hood because Joanne shares a relevant context to Faye's. Faye has acquired a true belief that there is a tree house at the edge of the forest and what Faye infers from this bit of knowledge is that Bronwyn's tree house is at the edge of the forest.

The tree house that Faye now knows is at the edge of the forest is nowhere near where Bronwyn's tree house ever would have been. Had Faye asked nearly *anybody else* in town, she would have been met with the answer, "No, there is no tree house at the edge of the forest." The just-as-likely conversation would have resulted in what would have counted for Faye as a false belief, but would have enabled her to transfer knowledge to Bronwyn, that there is no tree house at the edge of the forest, and thus that Bronwyn's tree house is gone. This bizarre chain of knowledge transferring through testimony is the result of the defeater for Faye that it just as easily could have been the case that the testimony she received had come from somebody who meant something different by 'forest.'

If this is plausible then it looks like Faye has no way of excluding the possibility that someone asserting *s* could just as easily have meant something else. If that's right, then even if a speaker *does* mean by *s* what is semantically uniform to Faye's utterance of 'forest,' the relevant defeater indicates that it would be fortuitous. Taken this way, a linguistic defeater for knowledge from testimony should give rise to new worries about the possibility of being Gettier'd on the testimonial transmission of knowledge.⁵

⁵ I'd like to thank audiences at Northwestern University (In particular, Ezra Cook, with whom I had several conversations regarding the project) and the University of Texas. Also, I would like to thank Jeremy Fantl who was my advisor at the University of Calgary when I began this project, with whom I developed the base cases for the problem.

REVIEWS

Mark R. Wynn, *Faith and Place. An Essay in Embodied Religious Epistemology*
(Oxford: Oxford University Press, 2009)

Reviewed by Ioan Alexandru Tofan¹

Mark Wynn's book (currently senior lecturer at Exeter University, UK) represents an original attempt to bring forward for discussion some of the classic paradigms of both theology and philosophy. The title itself is a surprising one (*Faith and Place. An Essay in Embodied Religious Epistemology*) and clearly specifies the theme set for research: the re-evaluation of the signification that 'the Place' has, and a new discussion about what *location*, as a phenomenological 'gesture' in relation with the religious experience, is.

The original place that Wynn's paper occupies in the bibliography of the philosophy of religion is secured: usually, the common debates are done about the de-location and the deconstruction of the space, as a specific reaction to the world of a certain religious attitude. Moreover, theology either overlooks the problem of *res extensa* as belonging to sublunary and transient, or handles it, as shown by the author, from the point of view of the divine attribute of the omnipresence. Therefore, either space is accidental, or formally created, and therefore locating a thing, phenomenon and, moreover, identifying the places as such, receives a decisive secondarity from the point of view of the theological reflection or of the religious experience. This type of metaphysical attitude is questioned by Wynn. From the point of view of the religious attitude as such, space receives signification and even an essential signification. It becomes *different*, it includes qualitative delimitations and, thus, it allows for the 'places' invested with meaning to occur. The experience of pilgrimage is suggestive, next to the religious gesture of space consecration, carefully and

¹ ACKNOWLEDGEMENT: This paper was made within *The Knowledge Based Society Project* supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815.

highly informed, among others, by Mircea Eliade. The very connection between the religious feeling and the characteristics of places with which it is connected—with the peace, or, on the contrary, with the tumult of world—may also be invoked. Finally, the scriptural meaning of ‘the high place’ or of the wilderness proves how spatiality is certainly not accidental in the context of religious experience. The information provided by the sociology, history or anthropology of religion are not however decisive for Wynn. The originality of the research lays in its phenomenological approach (actually quite rarely called as such) and in the attempt to re-enroll the religious experience in the specific ground of the inter-subjectivity, of the ‘world of life,’ where spatiality and location, as differentiated attitudes, significant in relation to space, become decisive.

An initial intention of the book is, therefore, to re-discuss the theological reflection regarding the divine omnipresence starting from the concrete experience of the religious man, the one rooted in a complex feeling of space and in an essential logic of location: “I am interested to see how the doctrine of divine omnipresence might be set within some larger theoretical framework which gives proper recognition (one that is neither psychologically reductionist nor metaphysically overbearing) to the place-relative character of religious belief and practice.” (pp. 4-5.) Wynn’s intention is not without precedent in the history of philosophy of religion. It witnesses an important moment through the writings of Hegel who, in an entirely different context, of course, also observed the need to double the theological reflection with the one regarding the concrete ways of religiosity, the only ones where the concept of divine being may ‘realize’ itself. Wynn’s phenomenological perspective resets the issue in entirely different frames and the intention to follow this issue starting from the signification which the religious man gives to the space and to the gesture of location is an original one.

The book’s second dimension is the epistemological approach of this issue. This time as well, the approach is an original one, although the sources named by the author are multiple. Wynn starts from the classical scenarios of analytical epistemology of religion—Richard Swinburne or William Alston. They act, most of the times, through the identification of some secular analogies of religious knowledge, whether scientific or perceptual. It is insufficient, argues Wynn. In fact, between the religious knowledge and the practical one there is a relation closer than the simple analogy. One of the possible connections may be obtained through ‘the knowledge of place’—

representing “an embodied, practical and, very often, theoretically inarticulate responsiveness to a given region of space.” (p. 8.) Two structural elements of the knowledge of place indicate the *personal* way of religious knowledge: the practical character and the super-individual, non-ostensive status of the object. Wynn’s presupposition is clearly expressed: religious knowledge is not a strictly intellectual one, it does not represent an exemplary form of *theory*, but it is both a cognitive positioning, as well as affective and practical, of the subject in relation to the divine being. This presupposition is not the only one in the history of philosophy. Augustine had already drawn attention on the fact that faith is uncertain, in the absence of the reference to the other Pauline virtues, hope and love. Faith, shows the Bishop of Hippo, does not imply a simple intellectual adherence, but a conversion of the soul in his reporting to the Boundlessness. The same (but this time from a phenomenological perspective, not theological-exegetical) is also for Wynn, the religious knowledge: a knowledge that belongs to the concrete, historical self, integrated in the world of life. Spatiality and the reference of the subject to the qualitative difference of places are structural elements of this world and, thus, they maintain an essential relation with religious knowledge. Wynn’s thesis from 2009 continues, in fact, a series of previous researches, mentioned by the author: *God and Goodness: A Natural Theological Perspective* (1999) and *Emotional Experience and Religious Understanding* (2005). The analogy invoked by Wynn, between *knowledge of place* and *religious knowledge*, has more than a rhetoric role. The two of them do not meet only under the structural aspect, but, in a fundamental manner, inside the life’s world, as concrete reports of subject not to objects as such, but to horizons and visibility conditions of objects.

The book’s chapters describe this original scenario, of the connection between location and religious experience. The two mentioned routes are followed: on the one hand, the relation between the concept of divine being and the concept of ‘place’ (Place) and, on the other hand, the structural connections between the two types of knowledge. The author also sketches, in the last chapters of the book, a series of considerations regarding a possible analysis of some cultural dimensions as a bridge between the two elements which face each other constantly in the book (for example, the aesthetic dimension, seen as a way of representation and identification of place, on the one hand, and as form of revelation, on the other hand).

The second chapter of the book is indeed interesting, taking the phenomenon of friendship as a model for the description of structure and of the signification of spatial knowledge (*knowledge of place*): “we have seen how the relationship of friends to a place may enable them to share various thoughts which it would be difficult to articulate in abstraction from the place. I have also noted that when friends reckon noninstrumentally with the character of a place, then their relationship to the place can come to assume, in some respects, the guise of a friendship.” (p. 43.) Friendship, claims Wynn with literary and biographical arguments, implies the emotional and intersubjective investment of reality. As such, it formally corresponds to the exercise of location, above defined. First of all, it sends to a concrete, integrated (*bodily*) way, of ‘response’ given to a place. Secondly, by means of the mechanism of emotional investment mentioned before, places become microcosms, they have the capacity to complete meanings that surpass them and represent *Totalities*. Thus becomes possible, as well, the integration of space into the religious experience. Both the consecration as religious act and the connection between *knowledge of place* and *religious knowledge* in the context of life’s world are based on the capacity of the *place* to function as a microcosm. The individual, daily or exceptional experience may thus, to the extent that it refers to a place, metaphysically open itself. Pilgrimage is the paradigmatic example in this case: “By means of the pilgrimage journey, the believer is able: (i) to encounter the significance of certain people or places (where this ‘encounter’ depends upon the believer placing themselves in a relevant relation of physical proximity); (ii) to achieve an embodied rather than purely mental or description-relative directedness to God; and (iii) to enact microcosmically, and so participate in, the Christian story—rather than simply professing that story in words.” (p. 172.)

In this context, the theological consequence of Wynn’s discourse becomes interesting. Namely, the analogy between the concept of space and the one of divine being does not explain only how it is possible for a place to obtain religious relevance. And vice versa, it founds a certain way of the divine, where this one can express cosmically and may be ‘recognized’ in different spatial situations. An entire series of issues is opened—and this is one of the important merits of the text submitted to debate: the problem of idolatry, for example, bears a philosophical translation starting from the concept of microcosm, the same as the abstract concept of divine being is, by means of the reinstatement of religious knowledge in life’s world, opened

towards an universe of significations and towards possible concrete representations. Wynn's choice is, therefore, to look at the religious experience not from the point of view of the divine omnipresence, but investigate the ways in which the qualitative differentiation of *places* inside this experience is established.

The sources of the discourse that I briefly presented here are multiple. First of all, it is about personal experience, about the friendship of the British poet Edmund Cusick. The approach is an interesting one and a phenomenological one. The 'daily,' the cotidianity, is for Wynn a source of significance and, at the same time, a methodological 'guide.' Cusick's poetic work counts as well as source for Wynn's reflection: "For Cusick of course, divine meanings are relative not just to bodily movements, but to places. His religious quest is, then, a search for the meanings which are inscribed in particular microcosmically significant places. And as his poetry reveals, these meanings may in turn be given in the storied identity of a place, or in the mode of bodily appropriation which it affords." (p. 248.) There are also mentioned theological sources, such as the paper of Friedrich von Hügel, *The Mystical Element of Religion as Studied in Saint Catherine of Genoa and Her Friends* (1923), from which Wynn keeps in mind the way in which theological knowledge is primarily oriented towards action and has an ethical relevance. Historical artifacts or significant places are essentially connected to religious experience. Susan White, David Brown and John Inge are also mentioned. Sometimes, the statements seem bold, from the point of view of theology as such ("the meaning of the world is God," p. 249). But, many times, they can be explained through an implicit philosophical reference in the text. The third 'source' of the text is the phenomenological one, visible especially through the way in which the two types of experience, the one of location and the religious one, are debated. Gerardus Van der Leeuw, Rudolf Otto and Mircea Eliade are quoted next to Maurice Merleau-Ponty. The phenomenological perspective is not assumed; it rather offers a language than a method. The book's intention, however, is not one of managing a philosophical analysis, but of formulating a series of reflections in a concrete manner, by crossing several theoretical discourses with the analysis of daily situations. And these reflections regard an interesting theme: recovering the categories of externality from the point of view of the religious experience, traditionally concentrated on the phenomenon of interiority. Abundant in examples and analogies, the text suggests, as mentioned, several openings and applications,

both in the field of esthetics or, broadly, of the philosophy of culture, as well as in the field of a 'hermeneutics' of the spatial dimension of the cotidianity.

Finally, the subtitle of the book sends, as well, to another direction of reading that may be developed. It is about identifying those ways of knowledge which cannot enter the abstract scheme of the relation subject-object. The religious knowledge or the one of place (*knowledge of place*) are epistemic ways deeply rooted in the world of life and in a series of non-theoretical relations, which describe horizons of the experience and not objects of it. Wynn's attempt of understanding the attribute of the divine omnipresence starting from the differentiated religious experience is crucial in this respect.

Cory Juhl, Eric Loomis, *Analyticity*
(New York: Routledge, 2010).

Reviewed by Viorel Țuțui¹

Since it was introduced by Immanuel Kant in the *Critique of Pure Reason*, the *analytic-synthetic* distinction had an intricate historical development. In some stages of this development it was considered to be one of the central problems in philosophical controversies, while in others it was virtually forgotten. In contemporary philosophy we witness a significant revival of this theme, beginning with the final decade of the last century. This resumption of the philosophical debate concerning the distinction between the *truths in virtue of meaning* and *truths in virtue of facts* occurred after a long interval in which it was practically ignored, because Quine's criticism against analyticity was considered compelling by the majority of contemporary philosophers.

In their recent work, *Analyticity*, Cory Juhl and Eric Loomis try to offer an introduction to the problem of the *analytic-synthetic* distinction. It represents both a historical and systematical overview of the problems concerning analyticity. The book is structured in six chapters: "Conceptions of Analytical Truth," "Carnap and Quine," "Analyticity and its Discontents," "Analyticity and Ontology," "Analyticity and Epistemology," and "Analyticity Reconsidered."

In the first chapter they present the historical emergence of the distinction beginning with what they call the 'prototype' of the *analytic-synthetic* distinction that was formulated by David Hume: the distinction between 'relations of ideas' and 'matters of fact.' They rightfully underline the fact that Hume's real interest was with the matters of fact and he didn't paid much attention to the relations of ideas. Kant was the philosopher that offered this distinction its philosophical significance in the context of distinguishing *empirical* from *necessary* truths. According to Kant's view, *analytic truths* are those judgments in which the subject A already contains the predicate B and the *synthetic judgments* are those in which

¹ ACKNOWLEDGEMENT: This paper was made within The Knowledge Based Society Project supported by the Sectoral Operational Programme Human Resources Development (SOP HRD), financed from the European Social Fund and by the Romanian Government under the contract number POSDRU ID 56815.

the predicate B lies outside the subject A, although it stands in connection with it. However, the Kantian theory regarding analyticity left many open questions. Their presentation continues with the theory of Bolzano, that extended the class of analytic truth to include logical truths and the theories of Frege and Russell that developed this contribution into the project of logicism, according to which all arithmetic truths can be derived from logical truths, and therefore they are analytical.

Juhl and Loomis affirm that this project was further developed by the Vienna Circle members, like Moritz Schlick and Rudolf Carnap, into a new and influential view that regarded *analytical truths* as expressions of the conventions governing language. (pp. 20-21.) Logical empiricists included in this category all logical and mathematical statements that were considered to be formal truths and opposed to empirical (factual) truths. In their view, any system of truths contains formal truths, that are created by stipulation and are governed only by consistency constraints, and empirical truths that connect the system with the empirical world. The final section of the first chapter briefly presents the main objective of the doctrine held by Rudolf Carnap: to provide a very general theory of objects and concepts that are conceived as logically constructed from the sensation language, understanding logic as a formal language based on conventions concerning the use of symbols. (p. 24.)

The second chapter opens with the analysis of the controversy between Rudolf Carnap and Willard Van Orman Quine concerning the *analytic-synthetic* distinction, which extends in the third chapter. The aim of this analysis is to prove that Quine's arguments fail to show that there is no philosophically interesting notion of analyticity. With this aim they reconstruct in the second chapter, in great detail, Carnap's broader philosophical project that he developed after the dissolution of the Vienna Circle. Responding to Gödel's discovery of the incompleteness of axioms systems for arithmetic, Carnap revised his project, starting with the book *The Logical Syntax of Language*, by assuming the plurality of logic, the liberty to construct logical systems and a language-relative conception of truth. In this context he redefined *analytical truths* as those statements which hold true solely in virtue of the rules of a language system. (p. 40.)

Juhl and Loomis underline the fact that, influenced by Tarski, Carnap tried to extend this project in order to replace troublesome philosophical concepts like 'reference' and 'truth' with more precise concepts of artificial languages. He distinguished the 'internal' question regarding what is truth in such a language from 'external' question about the utility of such a language as a whole. In this context, he attempts to 'explicate' the notion of analyticity as an internal concept

of particular artificial languages. The two authors mention the fact that Quine was influenced by the theory of Carnap regarding the 'explication' of philosophical concepts in more precise terms, but he modified this idea in the form of the 'elimination' thesis: we should eliminate all those concepts that prevent us to understand the world and its contents as physical phenomena. He develops this idea in the theory of 'radical translation,' according to which a language is nothing more than a physical phenomenon, a stimulus from which we cannot hypostatize linguistic phenomena like meanings and synonymies. From this perspective he argues that the notion of analyticity should be abandoned, insofar as it was explained in such terms.

In the first part of the third chapter, Juhl and Loomis present the development of Quine's critique of analyticity from the first expression it took in the article "Truth by Convention," in 1936, to its classical expression from the paper "Two Dogmas of Empiricism," in 1953, and ending with the latter form this criticism takes in the replies addressed to his critics that were included in the volume *The Philosophy of W.V.O. Quine* in 1986. They systemize all Quine's objections against analyticity in several categories: those which claim that 'analytic' is unintelligible, those which claim that analyticity is intelligible, but there are no instances in fact, and those that claim it is intelligible, but explanatorily useless. From Quine's critique of the first dogma (that of analyticity), they mention the arguments regarding the circularity of all the attempts to explain analyticity by using concepts like synonymy or meaning and his arguments regarding the fact that other concepts used to explain analyticity in formal languages, like that of 'semantic rule,' are equally problematic. From his critique against the second dogma (that of reductionism) they present the thesis of confirmational holism, according to which the language of science confronts with experience as a whole rather than confronting it sentence by sentence. All the statements within this comprehensive network are revisable in the light of experience. So, he leaves no place for *truths in virtue of meaning*.

Quine allowed that there could be 'legislative' definitions that could be conventionally true, but he denied that their status as conventional truths had any enduring consequences for the use of such sentences beyond the initial act of definition. However, Juhl and Loomis argue that he fails to acknowledge the important difference between two kinds of stipulation: the stipulation of a rule that prescribes the fact that something will hold and the hypothesizing that something will hold. It's a distinction between following a rule and merely engaging in some regular pattern of behavior. (pp. 121-123.) They will develop this distinction in the sixth chapter in a positive account of analyticity.

In the final part of the chapter, they emphasize the fact that, in his latter works, Quine modified his attitude regarding analyticity, allowing the fact that some notion of 'analytic,' understood as 'stimulus analyticity' and even as 'truth in virtue of meaning,' could be coherent, but he still denied that such a notion could possess any explanatory significance.

The forth chapter is dedicated to the ontological dispute that corresponds to the controversy regarding analyticity. It begins with the presentation of Quine's ontology, his physicalism, and of the relation between this ontological approach and the rejection of the intensional notions and entities that lead him to his 'nonfactualism' about reference and meaning. Juhl and Loomis consider his thesis of indeterminacy of translation as unsatisfactory. They mention Chomsky's objection according to which Quine uses a double standard in his attitude regarding physics and linguistics: he assumes the fact that the underdetermination of theory by the available data is problematic in the case of language, but not in the case of physics. (p. 148.) Next, they analyze the ontological approach defended by Carnap, and especially his distinction between internal and external problems. They argue that it faces many objections that seem to confirm Quine's thesis that the ontological questions (like that concerning the existence of numbers) cannot be settled only internally. In the final part of the chapter they mention the recent ontological contributions of Stephen Shiffer, Jody Azzouni, Eli Hirsh, Ted Sider and of the 'Canberra project' that seems to reproduce the controversy between Carnap and Quine in a contemporary context.

In the fifth chapter, their objective is to underline the philosophical debate regarding the connection between the epistemological problem of non-empirical truths and the problem of analyticity. They start by presenting what they called the 'classical position' regarding this connection that was developed by A.J. Ayer and was accepted by many other logical empiricists. In their opinion, this view is problematic because it conflates modal necessity, apriority, and generality and doesn't distinguish propositions from sentences. However, when Juhl and Loomis discuss Laurence Bonjour's objections against the conception of analyticity based on implicit definitions, although they accept that some of them are correct, they argue that these objections do not demonstrate the fact that any appeal to implicit definition is useless. And, after they expose some difficulties that Quine's approach faces, regarding the relation between analyticity and epistemology, they investigate the classical objection raised by Saul Kripke against the confusing of analyticity with apriority: his arguments for the existence of *a priori* and yet *contingent truths* and of *a posteriori necessary truths*. They try to reveal the fact

that some of his examples of *a priori contingent truths* are not genuine because there is not a single entity that is both known *a priori* and *contingent*. (207-208.)

The sixth and final chapter is the one that offers their positive conception concerning the problem of analyticity. They begin by analyzing what they consider to be the best cases for any advocate of analyticity, cases that are not affected by the objections raised by Willard Van Quine and Gilbert Harman: explicit stipulative definitions and mathematical stipulations. Extending the features of these paradigmatic examples, they underline the fact that analytic statements should be understood as true and empirically indefeasible, if we are ruling out empirical data regarding language use itself or second-order empirical data regarding the existence of non-empirical justification of mathematical truths. The starting point of their account of analyticity is the distinction between explicit stipulations and ‘hypotheses.’ They argue that many of the objections against the analyticity mentioned by Quine and Harman fail to acknowledge this distinction.

Another important conceptual difference they employ is that between *sentence*, understood as a linguistic expression, *statement*, which is a sentence together with some understood rules for using the sentence, and *propositions* that refer to the abstract objects which are correlated with the sentences. Using this distinction they define a special concept of analytic* that refers to statements as ‘sentences-as-used’: “When we introduce a stipulation of our particular indefeasible sort into our language, we introduce a coordinative rule concerning some stipulation sentence *s*, which states:

(Stip) Sentence *s* expresses some true proposition *p* (in our language *L*). Furthermore, the proposition *q*, that *s* expresses a true proposition (in *L*), is empirically indefeasible. No empirical evidence counts in favor of or against the truth of *q*.

When speakers of *L* accept *Stip* as a coordinative rule for speaking their language, we say that *s* is analytic* in *L*, or for speakers of *L*.” (pp. 218-219.)

They notice the fact that this sense of analytic* requires that *q* should be empirically indefeasible and doesn’t require that the proposition expressed by *s* should be empirically indefeasible too. This latter case is what they call ‘analytic-T’ (that is introduced by what they call a ‘transcendental stipulation.’) They use this difference to reject Kripke’s arguments for the existence of *contingent a priori truths*: his examples are analytic*, but not analytic-T: the propositions they express are empirically defeasible. Also, using the concept of analytic*, they reject a series of objections mentioned by Quine and Harman: the circularity objection, the indeterminacy of synonymy objection, the which/nonwhich objection. The sense

of analyticity they use does not require other concepts like necessity, synonymy, realism about meanings and so on. Statements that are analytic* are not true in virtue of their meanings: “what it is for a statement to be analytic* is to have the linguistic community take it as true and take it as indefeasible.” (p. 229.)

Another objection they analyze is the ‘non-explanatoriness’ objection: the fact that analyticity does not explain any empirical phenomenon. They accept this objection, but deny the fact that this makes the concept of analytic* useless: it can illuminate epistemologically puzzling phenomena such as our apparent non-empirical justification for believing some statements. The next objection they present is the classical ‘saying doesn’t make it true’ problem. Their response to this objection is that stipulations do not require any pragmatic or epistemic justification in order to be part of a coherent practice. In their opinion, a concept introduced by stipulation can be perfectly coherent, usable and comprehensible independently of whether there are any interesting applications for it and so independently of whether the stipulations involving it are empirically warranted. For example, they affirm that novel branches of mathematics generated by novel mathematical axiom systems or mathematical stipulations may have no known or expected non-mathematical uses. In their view, this thesis does not have the consequence that analytic statements (like mathematical statements) could not be applied in science since they might not be true or justified, because the practice of introducing these statements must be a coherent practice that has some rules of introduction that prevent difficulties like false empirical predictions: the introduction rules should not allow them. (pp. 231-232.)

Next, they argue that the concept of analytic* shouldn’t be confused with the traditional concept of analyticity: the class of analytic* statements doesn’t include empirical hypotheses and logical truths. This is the reason why their account of analyticity doesn’t solve all the epistemological problems that were traditionally connected with it. But they believe that it can be applied in some examples of apparently non-empirical knowledge and justification, like that of mathematical statements we mentioned above, and it can help us explain the status of these statements without any appeal to radical empiricism or *a priori* intuition. In their opinion, these statements are based on indefeasible stipulation. They think that, in this way, we can reject the objection according to which the mathematical statements are not arbitrary and therefore could not be analytic. A mathematical statement like “ $2+2=4$ ” is not arbitrary because that is what we mean by these mathematical concepts. Another difficulty they address is the objection which affirms that mathematical statements are empirically defeasible. They answer this objection by distinguishing the first-order canonical justification for

the mathematical proofs from the second-order justification for the assertion that there exists a first-order proof. Only the second-order justification is defeasible, but this doesn't affect the indefeasibility of mathematical statements themselves.

The most important difficulty for their account of mathematical statements is, in their opinion, the one that specifies the fact that existence claims cannot be stipulated. We can stipulate that, *if there are* mathematical objects, they will have some properties, but not *that there are* such mathematical objects. However, Juhl and Loomis affirm that their theory doesn't need to say that an act of mathematical stipulation created any entities, or caused the existential claim to be true: "What the stipulation brings about is facts concerning *what proposition a sentence expresses*, rather than *the truth of what is expressed*." (p. 253.) Moreover, they understand mathematical statements as expressing 'immune' propositions, propositions in the case of which no empirical proposition counts for or against. These propositions are about numbers, sets and other mathematical entities, and no empirical data could count against the existence of these entities. The last section of the book is dedicated to other potential applications of their account of analyticity.

In my opinion, the way they answer some of the objections against analyticity and especially those against the analyticity of mathematics is not satisfactory. I believe that they do not address the most important issue that the critics have in mind when they mention these objections: the problem of the objectivity of mathematical statements. The fact that, in their view, mathematical statements are based only on stipulations, transforms mathematics in a pure recreational game: the mathematicians just stipulate some rules of introduction for the mathematical concepts that are compatible with a coherent mathematical practice. But, this will make the applying of the mathematical statements in other sciences just a happy coincidence. However it will not explain the fact that virtually every mathematical topic proved to be very useful when it was applied in other scientific fields. Moreover, mathematical statements are not just useful for other sciences, like physics, but rather indispensable. We need a justification for the fact that this 'happy coincidence' always occurs, a justification Juhl and Loomis did not present.

A second problem is closely related to the first one: if Juhl and Loomis are right in affirming that we do not need a justification for our 'indefeasible' stipulations that are the basis of the mathematical statements, then what reasons do we have to consider that the applying of these statements in other sciences is justified? And, if the applying of mathematical statements in other sciences is not

justified, what reasons do we have to consider that the statements of these sciences themselves are justified?

But, if this is right, then the most important arguments they offer for the explanatory value of analyticity will not succeed. If analyticity* doesn't explain any of the epistemological problems that traditional analyticity was supposed to explain, then we will have good reasons for doubting its explanatory value and its usefulness.

Another observation we can make is the fact that their investigation of analyticity focuses mainly on the controversy between Quine and Carnap and less on the contemporary debate on this issue. Theories that are very important for this debate like those of Paul Boghossian, Christopher Peacocke, Paul Horwich, Bob Hale, Crispin Wright and Gillian Russell are briefly presented and some of them barely mentioned. I believe that a more extended analysis of the relation between their theory and this contemporary debate would have been clarifying and beneficial.

However, putting these difficulties aside, the book of Cory Juhl and Eric Loomis, *Analyticity*, is a very useful introduction to the problem of the philosophical significance of the *analytic-synthetic* distinction. One of its most important virtues is the fact that it offers a systematic investigation of the complex relations between analyticity and some of the important problems in the fields on epistemology and ontology, concerning the status and justification of mathematical sentences and the existence of mathematical objects.

NOTES ON THE CONTRIBUTORS

Fred Adams is Professor of Cognitive Science & Philosophy and Chair of the Department of Linguistics & Cognitive Science at the University of Delaware. His main areas of interest include: epistemology, cognitive science, philosophy of mind, ethics. He is the author of: *The Bounds of Cognition* (2008—with Kenneth Aizawa), *Ethical Issues in the Life Sciences* (2007) and *Ethical Issues for the 21 st Century* (2005). Contact: fa@udel.edu.

Scott F. Aikin is Senior Lecturer at the Philosophy Department of Vanderbilt University. His areas of interest are: epistemology, informal logic, argumentation, and pragmatism. His books include: *Epistemology and the Regress Problem* (2010), *Reasonable Atheism* (forthcoming 2011) and *Pragmatism: A Guide for the Perplexed* (2008)—both with Robert B. Talisse. Contact: scott.f.aikin@vanderbilt.edu.

John Barker is Associate Professor of Philosophy at the University of Illinois at Springfield, and Chair of the Philosophy Department. He is specialized in philosophical logic, philosophy of language, philosophy of mind and philosophy of science. His most recent published philosophical writings are “Disquotation, Conditionals, and the Liar” and “Undeniably Paradoxical” (both in *Polish Journal of Philosophy*). Contact: jbark3@uis.edu.

Vihren Bouzov is Associate Professor at the Faculty of Philosophy, University “St. Cyril and St. Methodius,” Bulgaria. His main interests are in: philosophy of science, logic of social sciences, philosophy of law. He is the author of: *The World of Norms* (2006), and *Logic of Decisions or How to be Effective in Decision Making* (2007), *Brief Notes on Philosophy* (2008)—in Bulgarian. Contact: v.bouzov@gmail.com.

James C. Cargile is Professor at the Department of Philosophy, University of Virginia and Director of the Colloquium Series. His areas of specialization are: metaphysics, philosophical logic, epistemology, ethics. He is the author of many articles, studies and chapters in books, such as: “The Fallacy of Epistemicism” (*Oxford Studies in Epistemology*, 2005), and “On Russell's Argument against

Resemblance Nominalism” (*Australasian Journal of Philosophy*, 2003). Contact: jc7y@virginia.edu.

Murray Clarke is a Professor of Philosophy at Concordia University, Montreal and Adjunct Research Professor at Carleton University's Institute of Cognitive Science. He works in cognitive science, philosophy of mind, and naturalized epistemology, and is the author of *Reconstructing Reason and Representation* (2004). He is currently working on a book defending an evolutionary account of informational atomism that combines a representational theory of mind with a computational account of cognition. Contact: murc@alcor.concordia.ca.

Sorin Costreie is Lecturer in the Department of Theoretical Philosophy at the University of Bucharest (UB). He is also here the Head of the Quality Management Department. His main philosophical interests are in the philosophy of mathematics and the philosophy of language. He published various papers related to these issues in collective volumes and specialized journals. He is also one of the co-organizers at UB of the *Bucharest colloquium in Analytic Philosophy* and the *Evenings with Science*. Contact: s.cos3e@yahoo.ca.

Theodore Everett is Professor of Philosophy at University SUNY Geneseo. His philosophical interests are in epistemology, metaphysics, logic, and ethics. He is the author of many studies and articles, including “Antiskeptical Conditionals” (*Philosophy and Phenomenological Research*, 2006), and “Analyticity Without Synonymy in Simple Comparative Logic” (*Synthese*, 2002), and “The Rationality of Science and the Rationality of Faith” (*Journal of Philosophy*, 2001). Contact: everett@geneseo.edu.

Carl Allen Ginet is Professor Emeritus at Sage School of Philosophy, Cornell University, Ithaca, New York. His work is primarily in action theory, moral responsibility, free will, and epistemology, but his interests include also metaphysics, philosophy of mind, and philosophy of language. Two of his most important contributions to philosophy are the books: *Knowledge, Perception, and Memory* (1975) and *On Action* (1990). Contact: cag2@cornell.edu.

Steven Hales is Professor at the Department of Philosophy, Bloomsburg University. His areas of specialization include epistemology and metaphysics. He is the author of *Relativism and the Foundations of Philosophy* (2006), *Nietzsche's Perspectivism* (2000—with Rex Welshon), and editor of many volumes such as: *A*

Companion to Relativism (forthcoming—2011) and *Analytic Philosophy: Classic Readings* (2001). Contact: hales@bloomu.edu.

Michael Harbour is a Ph.D. candidate in the Department of Philosophy at Vanderbilt University with a thesis on *Freedom as Anti-Domination*. His research focuses on social and political philosophy. He is the author of: “Nagel on Public Education and Intelligent Design” (*Journal of Philosophical Research*, 2010—with Scott F. Aikin and Robert B. Talisse) and “The Ethics of Inquiry and Engagement: The Case of Science in Public” (*Public Affairs Quarterly*, 2010—with Scott F. Aikin).

Lauren J. Leydon-Hardy is Teaching Assistant at Brandeis University, Massachusetts. Her area of specialization is epistemology, and other areas of competence are: philosophy of logic, philosophy of language and metaethics. She is currently working, at Brandeis University, on her MA thesis: *Justifying Falsehoods: False Justifiers and the Structure of Justification*. Contact: ljleydon@brandeis.edu.

Jonathan Neufeld is Assistant Professor of Philosophy and European Studies at Vanderbilt University. His research interests are in: philosophy of music, aesthetics, political philosophy, and philosophy of law. He is particularly interested in problems surrounding performance and interpretation. He is currently working on two books: *Music in Public: How Performance Shapes Democracy* (forthcoming 2011), and *Listeners, Critics, and Judges* (in preparation). Contact: jonathan.a.neufeld@vanderbilt.edu.

Robert B. Talisse is Professor of Philosophy and Political Science and Director of Graduate Studies in Philosophy at Vanderbilt University, and editor of *Public Affairs Quarterly*. His area of specialization is contemporary political philosophy. His most recent work engages issues at the intersection of political theory and epistemology. He is the author of: *Reasonable Atheism* (with Scott Aikin, forthcoming), *Democracy and Moral Conflict* (2009), *Pragmatism: A Guide for the Perplexed* (with Scott Aikin, 2008), *A Pragmatist Philosophy of Democracy* (2007), and *Democracy After Liberalism* (2005). He also edited many books on pragmatism, logic, political problems, American philosophy. Contact: robert.talisse@vanderbilt.edu.

LOGOS & EPISTEME

AIMS AND SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears in printed and on-line version in March, June, September, and December. Its fundamental mission is to support and publish various current reflections and researches that aim at investigating, analyzing, interpreting or philosophically explaining the human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal will publish 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 intends to promote 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.

NOTES TO CONTRIBUTORS

1. The journal accepts for publication articles, discussion notes and book reviews submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The editors of *Logos & Episteme* reserve the right to refuse all future papers belonging to the authors who fail to observe this rule.

2. 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.

3. 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.

4. 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.

5. 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.

6. 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.

7. 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.

8. A short CV including the author's affiliation and professional 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.

9. The reason for these requests is that all articles, with the exception of articles from the invited contributors, will be subject to a strict blind-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 review process is intended to take no more than six months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors. Processing of papers in languages other than English may take longer. 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.

10. 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.

11. 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. The papers containing racist and sexist opinions assumed by their authors will be rejected. The presence in texts of sexist or racist terms is accepted only if they occur in quotations or as examples.

12. 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.

13. Articles accepted for publication will become the property of *Logos & Episteme* and may not be reprinted or translated without the written permission of the editors. No manuscripts will be returned to their authors. The journal does not pay royalties. Authors of accepted manuscripts will receive free of charge two copies of the issue containing their papers.