PHILOSOPHICAL KNOWLEDGE

ITS POSSIBILITY AND SCOPE

Edited by

Christian Beyer

and

Alex Burri

PHILOSOPHICAL KNOWLEDGE

ITS POSSIBILITY AND SCOPE

Grazer Philosophische Studien

INTERNATIONALE ZEITSCHRIFT FÜR ANALYTISCHE PHILOSOPHIE

GEGRÜNDET VON
Rudolf Haller

HERAUSGEGEBEN VON
Johannes L. Brandl
Marian David
Leopold Stubenberg

VOL 74 - 2007

Rodopi

Amsterdam - New York, NY 2007

PHILOSOPHICAL KNOWLEDGE

ITS POSSIBILITY AND SCOPE

Edited by

CHRISTIAN BEYER

AND

ALEX BURRI

Die Herausgabe der GPS erfolgt mit Unterstützung des *Instituts für Philosophie der Universität Graz,* der *Forschungsstelle für Österreichische Philosophie, Graz*, und wird von folgenden Institutionen gefördert:

Bundesministerium für Bildung, Wissenschaft und Kultur, Wien
Abteilung für Wissenschaft und Forschung des Amtes der Steiermärkischen Landesregierung, Graz
Kulturreferat der Stadt Graz

In memoriam Georg Henrik von Wright

The paper on which this book is printed meets the requirements of "ISO 9706:1994, Information and documentation - Paper for documents - Requirements for permanence".

Lay out: Thomas Binder, Graz
ISBN: 978-90-420-2234-8
ISSN: 0165-9227
© Editions Rodopi B.V., Amsterdam - New York, NY 2007
Printed in The Netherlands

TABLE OF CONTENTS

PREFACE

The papers collected in this volume were written for the international conference *Philosophical Knowledge—Its Possibility and Scope* that took place from September 8th to September 10th 2005 in Erfurt. The aim of the conference was to bring together some of the protagonists of different metaphilosophical debates that have so far been led fairly independently of each other. In their contributions, the authors discuss the question of both the possibility and the scope of philosophical knowledge under a variety of aspects, particularly:

- a priori knowledge and the role of intuitions (A. Goldman, H. Kornblith, E. Sosa, T. Grundmann, T. Williamson),
- transcendental arguments (Q. Cassam, R. Stern, A. Burri),
- analytic philosophy and its methods (E. Brendel, M. Esfeld, H. Glock, F. Jackson, M. Willaschek), and
- phenomenology and analytic philosophy (D. Føllesdal, C. Beyer).

We should like to express our gratitude to the Fritz Thyssen Foundation for the generous grant that enabled us to organize the event in the first place; to the staff at the conference venue Begegnungsstätte Kleine Synagoge for their helpful support; to Silvère Schutkowski for his editorial assistance; to the German Research Foundation for a Heisenberg Scholarship which allowed one of the editors to prepare the volume for publication; to Marcus Willaschek for contributing a paper despite the fact that he had to decline our invitation to give a talk; and last, but by no means least, to all speakers for making the conference a success.

Christian BEYER
Alex BURRI

Grazer Philosophische Studien
74 (2007), 1–26.

PHILOSOPHICAL INTUITIONS: THEIR TARGET, THEIR SOURCE, AND THEIR EPISTEMIC STATUS

Alvin I. GOLDMAN
Rutgers University

Summary

Intuitions play a critical role in analytical philosophical activity. But do they qualify as genuine evidence for the sorts of conclusions philosophers seek? Skeptical arguments against intuitions are reviewed, and a variety of ways of trying to legitimate them are considered. A defense is offered of their evidential status by showing how their evidential status can be embedded in a naturalistic framework.

1. *Intuitions in philosophy*

One thing that distinguishes philosophical methodology from the methodology of the sciences is its extensive and avowed reliance on intuition. Especially when philosophers are engaged in philosophical "analysis", they often get preoccupied with intuitions. To decide what is knowledge, reference, identity, or causation (or what is the concept of knowledge, reference, identity, or causation), philosophers routinely consider actual and hypothetical examples and ask whether these examples provide instances of the target category or concept. People's mental responses to these examples are often called "intuitions", and these intuitions are treated as evidence for the correct answer. At a minimum, they are evidence for the examples' being instances or non-instances of knowledge, reference, causation, etc. Thus, intuitions play a particularly critical role in a certain sector of philosophical activity.

The evidential weight accorded to intuition is often very high, in both philosophical practice and philosophical reflection. Many philosophical discoveries, or putative discoveries, are predicated on the occurrence of certain widespread intuitions. It was a landmark discovery in analytic epistemology when Edmund Gettier (1963) showed that knowledge isn't equivalent to justified true belief. How did this "discovery" take place? It wasn't the mere publication of Gettier's two examples, or what he said about them. It was the fact that almost everybody who read Gettier's examples shared the *intuition* that these were not instances of knowing. Had their intuitions been different, there would have been no discovery. Appeals to intuition are not confined to epistemology; analytic philosophy as a whole is replete with such appeals. Saul Kripke remarks: "Of course, some philosophers think that something's having intuitive content is very inconclusive evidence in favor of it. I think it is very heavy evidence in favor of anything, myself. I really don't know, in a way, what more conclusive evidence one can have for anything, ultimately speaking" (1980: 42).

As a historical matter, philosophers haven't always described their methodology in the language of intuition. In fact, this seems to be a fairly recent bit of usage. Jaakko Hintikka (1999) traces the philosophical use of "intuition" to Chomsky's description of linguistics' methodology. In the history of philosophy, and even in the early years of analytic philosophy, the terminology of intuition is not to be found. Of course, historical philosophers dealt extensively with intuition in other contexts, but not in the context of appealing to particular examples and their classification. This is not to say that historical philosophers and earlier 20th-century philosophers did not make similar philosophical moves. They did make such moves, they just didn't use the term "intuition" to describe them. Consider Locke's presentation of the famous prince-cobbler case in his discussion of personal identity:

> For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon as deserted by his own soul, *every one sees* he would be the same person with the prince … (Locke 1694/1975: 44; emphasis added)

Locke says that every one "sees" that a certain classification—being the same as—is appropriate, and his term "sees" is readily translatable, in current terminology, as "intuits". Among ordinary-language philosophers of the mid-20th century, roughly the same idea was expressed in terms of what people would or wouldn't be inclined to *say*. One "would say" that the cobbler was the same person as the prince; one "wouldn't say" that a Gettier protagonist had knowledge. Here the propriety of saying or not saying something took the place of having an intuition; the matter was described in terms of speech inclinations rather than mental episodes. Nonetheless, the epistemological status of these inclinations or episodes played the same role in philosophical methodology. Each was invoked as a crucial bit of evidence for the philosophical "facts" in question.

2. *Skepticism about intuitions*

Nowadays philosophers routinely rely on intuitions to support or refute philosophical analyses, but a number of skeptics have emerged who raise challenges to this use of intuition. The skeptics include Robert Cummins (1998), Jonathan Weinberg, Shaun Nichols and Stephen Stich (2001), and Michael Devitt (1994). They dispute the evidential credentials or probity of intuitions. They deny that intuitions confer the kind of evidential support that they are widely taken to confer.

The grounds for skepticism are somewhat variable, but mostly they concern the fallibility or unreliability of intuitions, either intuitions in general or philosophical intuitions in particular. Here are three specific criticisms.

(1) Garden-variety intuitions are highly fallible. Why should philosophical intuitions be any different? If the latter are highly fallible, however, they shouldn't be trusted as evidence.

Garden-variety intuitions include premonitions about future events, intuitions about a person's character (based on his appearance, or a brief snatch of conversation), and intuitions about probabilistic relationships. These are all quite prone to error. What reason is there to think that philosophical intuitions are more reliable?

(2) People often have conflicting intuitions about philosophical cases. One person intuits that case x is an instance of property (or concept) F while another person intuits that case x isn't an instance of property (or concept) F. When such conflicts occur, one of the intuitions must be wrong. If the conflicts are frequent, the percentage of erroneous intuitions must be substantial and the percentage of correct intuitions not so high. Thus, the modest level of reliability of philosophical intuitions doesn't warrant assigning them significant evidential weight.

A third ground for skepticism doesn't appeal directly to the unreliability of intuition but rather to our inability to (independently) *know* or *determine* its reliability.

(3) The outputs of an instrument, procedure, or method constitute data we can properly treat as evidence only when that instrument, procedure, or method has been calibrated (Cummins 1998). Calibration requires corroboration by an independent procedure. Has intuition been calibrated? Has it been shown to be reliable by a method independent of intuition itself? There is no way to do this. Suppose we have a philosophical interest in fairness, and we ask people for their intuitions about the fairness of distributions described in certain hypothetical cases. We shouldn't trust their intuitions about these cases unless we have antecedently determined that their fairness intuitor is reliable, i.e., unless it has been calibrated. But how can we perform this calibration? We don't have a "key" by which to determine which outputs of their intuitor are correct, and there is no key to be found.

3. *Initial responses to skeptical challenges*

For each of these skeptical challenges, there appear to be at least initially plausible responses. In response to challenge (1), a defender of philosophical intuition would want to distinguish between different types of intuitions. First, the intuitions we have here identified are what might be called *classification* or *application* intuitions, because they are intuitions about how cases are to be classified, or whether various categories or concepts apply to selected cases.[1] This in itself, however, provides no reason for thinking that philosophical intuitions are epistemically superior to garden-variety intuitions. Why should classification or application intuitions be superior? A supplementary response is that application intuitions are a species of *rational* intuitions, and that rational intuitions are more reliable than others. Many authors are sympathetic to this approach, but George Bealer (1998) has been most forceful in championing it. Bealer distinguishes between physical and rational intuitions, and regards only the latter as having special epistemic worth. We shall return to this below.

[1] Frank Jackson (1998) also views classification, or application, intuitions as the central type of philosophical intuition.

In response to challenge (2), a defender of philosophical intuitions might urge caution. It remains to be seen just how extensive are the conflicts in application intuitions across different individuals. Moreover, whether the conflicts are genuine depends on the precise contents of the intuitions, or what they are taken to be evidence *for*. It is possible that a state of affairs for which one person's intuition is evidence doesn't really conflict with a state of affairs for which another person's intuition is evidence, even when there is a "surface" conflict. I'll return to this point below as well.

In response to challenge (3), a defender of philosophical intuitions might reject Cummins's epistemological presuppositions. The defender might say that independent corroboration, or calibration, of an instrument, procedure, or method is too stringent a requirement on its evidence-conferring power. In particular, there must be some procedures or methods that are *basic*. In other terminology, there must be some basic "sources" of evidence. Basic sources are likely to include mental faculties such as perception, memory, introspection, deductive reasoning, and inductive reasoning. These faculties are all regarded, by many or most epistemologists, as bona fide sources of evidence. Yet all or many of these sources may be basic in precisely the sense that we have no independent faculty or method by which to establish their reliability. Yet that doesn't undercut their evidence-conferring power. Consider memory, for example. Memory may be our basic way of forming true beliefs about the past. All other ways of gaining access to the past depend on memory, so they cannot provide *independent* ways of establishing memory's reliability (see Alston 1993). If we accept Cummins's constraint on evidencehood, the outputs of memory will not constitute legitimate data or pieces of evidence. But that is unacceptable, on pain of general skepticism. It is better to accept the conclusion that basic sources of evidence don't have to satisfy the calibration, or independent corroboration, constraint. Intuition may be among the basic sources of evidence.

Although Cummins's independent corroboration condition on a source of evidence is too stringent, it seems reasonable to substitute a weaker condition as a further requirement on evidencehood. This weaker condition is a "negative" one, viz., that we *not* be justified in believing that the putative source is *un*reliable. A possible variant is the condition that we *not* be justified in *strongly doubting* that the source is reliable. The latter negative condition will sometimes be invoked in the discussion to follow.

4. *The targets of philosophical analysis*

In response to skeptical challenge (2), I said that resolution of this challenge requires a more careful inquiry into the precise targets of philosophical analysis. Philosophical analysis, of course, doesn't simply aim to answer questions about particular cases. Epistemology isn't much interested in whether this or that example is an instance of knowledge; rather, it aims to say what knowledge is in general, or something in that ballpark. Individual cases are typically introduced as test cases of one or more general accounts. Depending on how a case is classified, it might falsify a general account or corroborate it. But what, exactly, does philosophical analysis aim to give general accounts *of*? Knowledge, causation, personal identity, and so forth are typical examples of categories that absorb philosophy, but different theorists have different conceptions (often unstated) of how, exactly, these targets are to be construed. A choice among these different construals can make a big difference to the viability of intuition as a source of evidence about the targets, because many construals invite *strong doubts* that the source is reliable. Let us examine five ways of construing the targets.

(1) Platonic forms
(2) Natural kinds
(3) Concepts$_1$—concepts in the Fregean sense
(4) Concepts$_2$—concepts in the psychological sense, specifically, the individualized, personal sense
(5) Concepts$_3$—shared concepts$_2$

The first two construals invoke entities that aren't described as concepts. Each is some sort of non-conceptual entity that exists entirely "outside the mind". According to the first construal, philosophy aims to obtain insight into (e.g.) the form of the Good, and other such eternal, non-spatially-located entities. According to the second construal, knowledge, causation, personal identity, and so forth are "natural" properties or relations, which exist and have their distinctive characteristics quite independently of anybody's concepts or conception of them, like water or electricity.

There are two questions to be posed for each of these (and similar) construals. First, under this construal how could it plausibly turn out that intuitions are good evidence for the "constitution" or characteristics of the targets? Second, does this construal comport with the actual intuitional methodology used by analytic philosophers? Start with construal (1). If the target of philosophical analysis is the constitution or composition of Platonic forms (or their ilk), the question is why an episode that occurs in somebody's mind—an episode of having an intuition—should count as evidence about the composition of a Platonic form.[2] If someone experiences an intuition that the protagonist in a selected Gettier example doesn't know the designated proposition, why should this intuitional experience be *evidence* that the form KNOWLEDGE is such that the imaginary protagonist's belief in this proposition doesn't "participate" in that form? What connection is there between the intuition episode and the properties of the form KNOWLEDGE such that the intuition episode is a reliable indicator of the properties of KNOWLEDGE? (I am assuming, for argument's sake, that this form exists.) We have reason to seriously *doubt* the existence of a reliable indicatorship relation.

Notice that it doesn't much matter how, exactly, we characterize intuitions. Whether intuitions are inclinations to believe, or a *sui generis* kind of seeming or propositional attitude (see Bealer 1998: 207), it is still a puzzle why the occurrence of such a mental event should provide evidence for the composition of a Platonic form. Compare this case with perceptual seemings and memory seemings. In these cases we know (in outline, if not in detail) the causal pathways by which the properties of an external stimulus can influence the properties of a visual or auditory experience. With this kind of dependency in place, it is highly plausible that variations in the experience reflect variations in the stimulus. So the specifics of the experience can plausibly be counted as evidence about the properties of the stimulus. Similarly in the case of memory, what is presently recalled varies (counterfactually) with what occurred earlier, so the specifics of the recall event can be a reliable indicator of the properties of the original occurrence. But is there a causal pathway or counterfactual dependence between Platonic forms and any mental "registration" of them? A causal pathway seems to be excluded, because Platonic forms are not spatio-temporal entities. A counterfactual dependence is not impossible, but there is reason to doubt that such a dependence obtains. I here register the general sorts of qualms that have long plagued traditional accounts of rational insight or "apprehension" of abstract entities. These accounts leave too many mysteries, mysteries that undercut any putative reliability needed to support a reflective acceptance of an evidential relation-

[2] For an earlier treatment of this question, and analogous questions for the other construals of the targets, see Goldman and Pust (1998).

ship between intuitional episodes and their targets construed as abstract entities.

Let us turn now to construal (2), the natural kinds construal, which has been formulated and championed by Hilary Kornblith (2002). Kornblith emphasizes that natural kinds are "in the world" phenomena, emphatically not merely concepts of ours. He rejects concepts as the objects of epistemological theorizing on the ground that by bringing concepts into an epistemological investigation, "we only succeed in changing the subject: instead of talking about knowledge, we end up talking about our concept of knowledge" (2002: 9–10). For Kornblith, the methodology of consulting intuitions (within epistemology) is part of a *scientific* inquiry into the nature of knowledge, closely akin, to use his example, to what a rock collector does when gathering samples of some interesting kind of stone for the purpose of figuring out what the samples have in common. Let us examine this approach.

Presumably, an inquiry into the composition of a natural kind is an inquiry into a *this*-world phenomenon. Even if natural kinds have the same essence or composition in every possible world in which they exist, the question for natural science is which of the conceivable natural kinds occupy *our* world. Does this feature of scientific inquiry into natural kinds mesh with the philosophical practice of consulting intuitions? No. A ubiquitous feature of philosophical practice is to consult intuitions about merely conceivable cases. Imaginary examples are treated with the same respect and importance as real examples. Cases from the actual world do not have superior evidential power as compared with hypothetical cases. How is this compatible with the notion that the target of philosophical inquiry is the composition of natural phenomena? If philosophers were really investigating what Kornblith specifies, would they treat conceivable and actual examples on a par? Scientists do nothing of the sort. They devote great time and labor into investigating actual-world objects; they construct expensive equipment to perform their investigations. If the job could be done as well by consulting intuitions about imaginary examples, why bother with all this expensive equipment and labor-intensive experiments? Evidently, unless philosophers are either grossly deluded or have a magical shortcut that has eluded scientists (neither of which is plausible), their philosophical inquiries must have a different type of target or subject-matter.

In responding to criticisms of this sort, Kornblith (2005) indicates that although he regards epistemology as an empirical discipline, it nonetheless investigates necessary truths about knowledge. Just as it is a necessary truth that water is H_2O, so there are various necessary truths about knowledge, and it is epistemology's job to discover these truths. Might this be why it is legitimate for epistemologists to adduce merely conceivable examples, involving other possible worlds? Kornblith doesn't say this, and it seems inadequate as a potential response. While it may be a necessary truth that water is H_2O, scientists first have to discover that what water is (in the actual world) is H_2O, and Kornblith admits that this must be an empirical discovery. Intuitive reactions to merely imaginary cases are not part of such an empirical procedure. Similarly, we cannot scientifically discover what knowledge is in the actual world by consulting intuitions about imaginary cases. So why do philosophers engage in this activity?

When I raise this point (Goldman 2005) in discussing Kornblith's book, he concedes that his approach doesn't explain philosophers' preoccupation with imaginary examples. He adds: "Goldman may have underestimated the extent to which I believe that standard philosophical practice should be modified" (2005: 428). So Kornblith agrees that, so long as we are discussing existing philosophical practice, his kind of naturalism cannot do the job. But he holds that existing practice is somehow inadequate or objectionable. I shall return to these concerns of his at the end of this paper. For now I reiterate the point that as long as we are merely trying to describe or elucidate existing practice, the natural kinds approach (as Kornblith spells it out) cannot be right.

5. *Concepts in the Fregean sense*

We turn now to the third proposed construal, concepts in the Fregean sense of "concept", which we called "concepts_1". In this sense, concepts are abstract entities of some sort, graspable by multiple individuals. These entities are thought of as capable of becoming objects of a faculty of intuition, *rational* intuition. Moreover, philosophers like Bealer (1998) want to say that rational intuitions are sufficiently reliable to confer evidence on the appropriate classification (or "application") propositions. Indeed, rational intuition is a faculty or source that is *modally reliable* (in Bealer's terminology). Two questions arise here: What distinguishes rational intuitions from other types of intuition, and is there good reason to think that rational intuitions—specifically, the sub-category of classification intuitions—have the needed properties to qualify as an evidential source?

According to Bealer, rational intuitions are distinguishable from other (e.g., physical) intuitions in virtue of the fact that rational intuitions have a sort of modal content. "[W]hen we have a rational intuition—say, that if P then not not P—it presents itself as necessary; it does not seem to us that things could be otherwise; it must be that if P then not not P." (1998: 207) Bealer goes on to say that application intuitions, i.e., intuitions to the effect that a certain concept does or does not apply to a certain case, are a species of rational intuitions. He is not sure how to analyze what it means for an intuition to present itself as necessary (and hence to be a rational intuition), but offers the following tentative proposal: "necessarily, if x intuits that P, it seems to x that P and also that necessarily P" (1998: 207).

Does this work? How are we to understand the initial operator "necessarily"? Is it metaphysical necessity? So understood, the claim can't be right. It implies that it is metaphysically impossible for there to be any creature for whom it seems that 18 + 35 = 53 but for whom it doesn't seem that *necessarily*, 18 + 35 = 53. But such a creature surely is possible. For starters, there could be a creature that understands arithmetic but doesn't understand modality. Second, there could be a creature that understands both arithmetic and modality but forms intuitions about modality more slowly than intuitions about arithmetic. At some moments, it seems to this creature that the foregoing arithmetic sum is correct but it doesn't yet seem to him that it is necessary. The same point applies to application intuitions. Presented with a Gettier example, it strikes a beginning philosophy student that this is not an instance of knowing, but it doesn't strike the student as necessarily true. I suspect this is the actual condition of many beginning philosophy students. They have application intuitions without any accompanying modal intuitions.

A different approach to the explication of rational intuitions is pursued by Ernest Sosa (1998). In seeking to identify intuition in the philosophically relevant sense, Sosa places great weight on the content of an intuition being *abstract*. "To intuit is to believe an abstract proposition merely because one understands it and it is of a certain sort ..." (1998: 263–264). Should rational or intellectual intuitions be restricted to ones whose contents are abstract propositions? Sosa characterizes abstract propositions as ones that "abstract away from any mention of particulars" (1998: 258). But this definition threatens to exclude our primary philosophical examples, viz., application intuitions. These often concern particulars, both particular individuals and particular situations. Thus, Sosa's account threatens to rule out the very examples that most interest us.

If we can't unify rational intuitions in terms of their *contents*, perhaps they can be unified in terms of their *phenomenology*. Perhaps a common phenomenology unites intuitions concerning logic, mathematics, and conceptual relationships. What might this common phenomenology be? A phenomenological feature they share is the feeling that they come from "I know not where". Their origins are introspectively opaque. This isn't helpful, however, to rationalists of the type under discussion. *All* intuitions have this opaqueness-of-origin phenomenology, including garden-variety intuitions like baseless hunches and conjectures, which are rightly disparaged as unreliable and lacking in evidential worth. Grouping application intuitions with this larger, "trashy" set of intuitions is likely to contaminate them, not demonstrate their evidential respectability.

This problem might be averted if we turn from phenomenology to psychological origins, including unconscious psychological origins. Hunches and baseless conjectures presumably lack a provenance comparable to that of mathematical, logical, or application intuitions. So unconscious origin looks like a promising basis for contrasting these families of intuitions. There is a serious problem here, though. It is unlikely that there is a single psychological faculty responsible for all intellectual insight. The psychological pathways that lead to mathematical, logical, and application intuitions respectively are probably quite different. Elementary arithmetic intuitions, for example, are apparently the product of a domain-specific faculty of numerical cognition, one that has been intensively studied in recent cognitive science (Dehaene 1997). There is no reason to expect logical intuitions to be products of the same faculty. Application intuitions are likely to have still different psychological sources, to be explored below. So if the suggestion is that application intuitions should be grouped with mathematical and logical intuitions because of a uniform causal process or faculty of intellectual insight, this is psychologically untenable. It is initially plausible because they are not phenomenologically distinguishable. But if causal origin runs deeper than phenomenology—as it surely does—then the sameness-of-psychological-origin thesis is unsustainable. Moreover, difference of psychological origin is important, because it undercuts the notion that rational intuitions are homogeneous in their reliability. Arithmetic intuitions might be reliable—even modally reliable—without application intuitions being comparably reliable.

If the targets of application intuitions are Fregean concepts, does this inspire confidence that such intuitions are highly reliable? Oddly, Bealer himself makes no claim to this effect; his central claim is vastly more cautious. Bealer acknowledges that concepts can be possessed either weakly or strongly. Weak possession is compatible with misunderstanding or incomplete understanding. Only strong possession, which Bealer calls "determinate" concept possession, carries with it a guarantee of truth-tracking intuitions. However, Bealer offers no guarantee that either ordinary people or philosophers who possess a concept will possess it determinately. In the concluding section of his 1998 paper, Bealer summarizes his argument (in part) as follows: "With this informal characterization in view, intuitive considerations then led us to the *possibility of determinate possession*, the premise that it should be at least possible for most of the central concepts of philosophy to be possessed determinately" (1998: 231, emphasis in the original). If the determinate possession of philosophical concepts is merely *possible*, and by no means guaranteed or even probable, why should philosophers rely on ordinary people's intuitions as guides to a concept's contours? No evidence is provided that people, especially lay people, actually grasp selected philosophical concepts determinately. So Bealer's approach provides no solid underpinning for the philosophical practice of consulting ordinary people's application intuitions.

Finally, construing Fregean concepts as the targets of application intuitions doesn't safeguard against the possibility of different people having different application intuitions about the same concept and example. If there are many instances of such conflicts, these intuitions won't have even high *contingent* reliability, much less high *modal* reliability. Traditionally, philosophers haven't worried much about this prospect. But some of the intuition skeptics mentioned at the outset worry very much about it. Jonathan Weinberg, Shaun Nichols and Stephen Stich (2001) have done studies of people's intuitions, including intuitions about the applicability of the knowledge concept in Gettier-style cases. In contrast to the widespread view among epistemologists that Gettier-style cases prompt highly uniform intuitions, they found substantial divergences in intuition, surprisingly, along cultural lines. Undergraduate students at Rutgers University were used as subjects, and were divided into those with Western (i.e., European) ethnicities versus East Asian ethnicities. In one study involving a Gettier-style case, a large majority of Western subjects rendered the standard verdict that the protagonist in the example "only believes" the proposition, whereas a majority of East Asian subjects said the opposite, i.e., that the protagonist "knows" (2001: 443; see Figure 5). If cases like this are rampant (and that remains to be shown), it's a non-trivial challenge to the reliability of application intuitions under the Fregean concept construal.

6. *Concepts in the personal psychological sense*

Suppose that the target of philosophical analysis is concepts, but concepts in the psychological rather than the Fregean sense. In this sense, a concept is literally something in the head, for example, a mental representation of a category. If there is a language of thought, a concept might be a (semantically interpreted) word or phrase in the language of thought. What I mean by a *personal* psychological sense of concept is that the concept is fixed by what's in its owner's head rather than what's in the heads of other members of the community.[3] It's an individual affair rather than a social affair. This does not prejudice the case for a separate sense of "concept" pertaining to a community (what I mean to denote by "concept$_3$").

A chief attraction of construing concepts$_2$ as the targets of philosophical analysis (though perhaps not the ultimate targets) is that it nicely handles challenges to the reliability of intuition arising from variability or conflicts of intuitions across persons. If the targets are construed as concepts in the personal psychological sense, then Bernard's intuition that F applies to x is evidence only for *his* personal concept of F, and Elke's intuition that F doesn't apply to x is evidence only for *her* personal concept of F. If Bernard intuits that a specified example is an instance of knowledge and Elke intuits otherwise, the conflict between their intuitions can be minimized, because each bears evidentially on their own personal concepts, which may differ. This may be precisely what transpires in the cases reported by Weinberg et al. Under this construal of the evidential targets, interpersonal variation in intuitions doesn't pose a problem for intuitional reliability, because each person's intuition may correctly indicate something about

[3] This is not intended as a position statement on the wide/narrow issue concerning the contents of thought. It may be that thought contents *in general* do not supervene simply on events that transpire in an individual thinker's head. Nonetheless, the specific thoughts of each person—including the specific concepts each entertains—are a special function of what goes on in that individual's head rather than anybody else's. If Jones never entertains the thought that aardvarks drive automobiles, his never entertaining it is a function of what happens in his head rather than any other person's head. And if he never entertains the concept of an aardvark, this is a function of what happens in his head rather than any other person's head—at least of what happens in his head in interaction with the environment rather than what happens in any other person's head in interaction with the environment.

his or her concept$_2$, viz., whether the concept$_2$ does or doesn't apply to the chosen example.

It must be conceded that when a person thinks the thought, or has the intuition, "The Gettier disjunction case isn't an instance of knowledge", the content of the thought is not self-referential. It isn't naturally expressed as, "The Gettier disjunction case isn't an instance of my *personal* concept of knowledge". Nonetheless, epistemologists are at liberty to take the person's intuition, or thought, as evidence for a proposition concerning that person's individualized, psychological concept. This is what I propose to do.

But why is a person's intuition *evidence* for a personal psychological concept? I assume that any evidential relationship depends, at a minimum, on a relation of reliable indicatorship. But what makes such a relation hold in the case of application intuitions and concepts$_2$? Do we have reason for thinking that it holds? And do we avoid reasons for seriously *doubting* the existence of a reliable indicatorship relation?

Distinguish two approaches to the relation between concepts and evidencehood: *constitutive* and *non-constitutive* approaches. A constitutive approach can be illustrated by reference to phenomenalism (or other assorted versions of idealism). According to phenomenalism, what it *is* to be a physical object of a certain sort is that suitably situated subjects will experience perceptual appearances of an appropriate kind. Appearances of the appropriate kind are not only evidence for a physical object of the relevant sort being present, but the evidentiary relation is *constitutively* grounded. The evidentiary status of appearances is grounded in the very constitution of physical objects. Physical objects are precisely the sorts of things that give rise to appearances of the kind in question. According to realism, by contrast, to be a physical object has nothing essentially to do with perceptual experience. True, physical objects may cause perceptual experiences, but what they *are* (intrinsically) is wholly independent of perceptual experience. This view is compatible with perceptual experiences qualifying as evidence for the presence of appropriate physical objects, but here the evidential relation would not be constitutively grounded. There are many possible theories of non-constitutive evidencehood; I won't try to survey such theories. What is important for the moment is simply the distinction between constitutive and non-constitutive groundings of evidential relations.

Although I don't support phenomenalism, I am inclined to support a parallel theory for the evidential power of application intuitions. I think that the evidential status of application intuitions is of the constitutively-grounded variety. It's part of the nature of concepts (in the personal psychological sense) that possessing a concept tends to give rise to beliefs and intuitions that accord with the contents of the concept. If the content of someone's concept F implies that F does (doesn't) apply to example x, then that person is disposed to intuit that F applies (doesn't apply) to x when the issue is raised in his mind. Notice, I don't say that possessing a particular concept of knowledge makes one disposed to believe a correct *general* account of that knowledge concept. Correct general accounts are devilishly difficult to achieve, and few people try. All I am saying is that possessing a concept makes one disposed to have pro-intuitions toward correct applications and con-intuitions toward incorrect applications—correct, that is, relative to the contents of the concept as it exists in the subject's head. However, our description of these dispositions must be further qualified and constrained, to get matters right.

There are several ways in which application intuitions can go wrong. First, the subject may be misinformed or insufficiently informed about example x. Her intuitive judgment can go awry because of an erroneous belief about some detail of the example. Second, although she isn't misinformed about the example, she might forget or lose track of some features of the example while mentally computing the applicability of F to it. Third, the subject might have a false theory about her concept of F, and this theory may intrude when forming an application intuition. It's important here to distinguish between a theory presupposed by a concept and a theory *about* the concept, i.e., a general account of the concept's content. Here I advert only to the latter. Any of these misadventures can produce an inaccurate intuition, i.e., inaccurate relative to the user's own personal concept. For these reasons, intuitions are not infallible evidence about that personal concept.

These points go some distance toward explaining actual philosophical practice. First, philosophers are leery about trusting the intuitions of other philosophical analysts who have promoted general accounts of the analysandum, e.g., knowledge or justification. Commitment to their own favored account can distort their intuitions, even with respect to their own (pre-theoretical) concept. Second, because erroneous beliefs about an example can breed incorrect intuitions, philosophers prefer stipulated examples to live examples for purposes of hypothesis testing. In a stipulated example, the crucial characteristics of the example are highlighted for the subject, to focus attention on what is relevant to the general account currently being tested.

Although errors in application intuitions are possible, a person's application intuitions vis-à-vis their own personal concepts are highly likely to be correct if the foregoing safeguards are in place. Thus, the reliability criterion for evidence-conferring power—one very natural criterion (or partial criterion)—is met under the concepts$_2$ construal of the targets of philosophical analysis.

Another virtue of the concepts$_2$ approach is the congenial naturalistic framework it provides for the respectability of application intuitions as evidence. Unlike Platonic forms, natural kinds, or Fregean concepts, there can be a clear *causal* relationship between personal concepts and application intuitions concerning those concepts. Although psychological details remain to be filled in, there is nothing inherently mysterious in there being a causal pathway from personal psychological concepts to application intuitions pertaining to those concepts. Personal psychological concepts can be expected to produce accurate intuitions concerning their applicability. So as long as the various threats of error of the kinds enumerated above aren't too serious, high reliability among application intuitions is unperplexing and unremarkable under the concepts$_2$ approach. Although naturalistically-minded philosophers are understandably suspicious and skeptical about intuitions and their evidential *bona fides*, here we have a satisfying resolution to the challenge from naturalistic quarters, a resolution that copes straightforwardly with existing evidence of interpersonal variation in intuitions. Thus, I share with Kornblith the aim of obtaining an epistemology of philosophical method that sits comfortably within a naturalistic perspective. Whereas Kornblith's naturalism leads him to extra-psychological objects as the targets of philosophical theory and to very limited acceptance of intuitional methodology, my psychologistic brand of naturalism leads to personal psychological concepts as the initial targets of philosophical analysis and to a greater acceptance of standard intuitional methodology.

7. *Shared and socially fixed concepts*

A predictable response to our proposal is that even if intuitions constitute evidence for personal psychological concepts, that's not a very interesting fact. Personal concepts can't be all—or even very much—of what philosophy is after. Fair enough. I am not saying that the analysis of personal concepts is the be-all and end-all of philosophy, even the analytical part of philosophy. But perhaps we can move from concepts$_2$ to concepts$_3$, i.e., shared (psychological) concepts. This can be done if a substantial agreement is found across many individuals' concepts$_2$. Such sharing cannot be assumed at the outset, however; it must be established. Philosophers often presume that if their own and their colleagues' intuitions point to a certain conclusion about a concept, that's all the evidence needed. If discerning judges agree in matters of concept application, then other judges would make the same assessment. The empirical work of Weinberg, Nichols and Stich (2001), however, raises doubts about this. And we all know from even casual philosophical discussion that philosophers don't always share one another's intuitions. Moreover, intuitive disagreement is probably underreported in the literature, because when philosophers publish their work they typically avoid examples they know have elicited conflicting intuitions among their colleagues. So the extent of disputed intuitions may be greater than philosophers officially acknowledge, and this may challenge the hope of identifying unique, socially shared concepts.

To safeguard some sort of supra-individual conception of concepts, there are other ways to proceed. One possibility is not to place the personal concepts of all individuals on a par, but to privilege some of them. How might this be done? There are several possibilities, some appealing to metaphysics and some to language. An appeal to metaphysics might return us to the natural kinds approach. Concepts that correspond to natural kinds should be privileged, those that don't, shouldn't. The problem here is that it's doubtful that every target of philosophical analysis has a corresponding natural kind. Take knowledge again as an example. A popular view in contemporary epistemology (with which I have much sympathy) is that knowledge has an important context-sensitive dimension. The exact standard for knowledge varies from context to context. Since it seems unlikely that natural kinds have contextually variable dimensions, this renders it dubious that any natural kind corresponds to one of our ordinary concepts of knowledge.

A more promising approach is to recast the entire discussion in terms of language. Concepts are the meanings of (predicative) words or phrases (Jackson 1998: 33–34). The correct public concept of knowledge is the meaning of "know". Many people who use the word "know" and its cognates may not have a full or accurate grasp of its meaning. Their intuitions should be ignored or marginalized when we try to fix the extension and intension of the term. Only *expert* intuitions should be consulted. This is a natural line of development of Putnam's (1975) theme that meanings are determined by a division of linguistic labor in which experts play a central role.[4]

I hesitate to go down this road for two reasons. First, the idea of a division of linguistic labor, in which deference to linguistic experts holds sway, makes most sense for technical terms that aren't mastered by ordinary users of the language. Clearly, it would be a mistake for philosophical theorists to rely on the classification intuitions of users with inadequate mastery of the meanings of the words. However, concepts expressed by technical terms are not the chief concern of philosophical analysis. Philosophical analysis is mainly interested in common concepts, ones that underpin our folk metaphysics, our folk epistemology, our folk ethics, and so forth. I don't say this is *all* that philosophy is or should be concerned with. But when philosophers engage in analysis, folk concepts are what preoccupy them (Jackson, 1998). In this terrain, there isn't any significant expert/novice divide. Thus, if there are still differences in personal concepts associated with a single word, the differences cannot be resolved by appeal to (semantic) experts.

Second, there is a general problem with any attempt to configure the conceptual analysis enterprise in purely linguistic terms. Many of our most important folk-ontological concepts, I submit, are prior to and below the level of natural language. For instance, our unity criteria for physical objects fix the contours of single whole objects without recourse to predicates of natural language. They are independent of particular linguistic sortals, as illustrated by our ability to visually pick out a unitary physical object without yet deciding what *kind* of object it is. ("It's a bird, it's a plane, no, it's Superman!") Indeed, deployment of such criteria is a prerequisite for children to acquire mastery of verbal sortals. Children must already pick out unified physical objects in order to learn (at least with approximate accuracy) what adults refer to by such sortals as "rabbit", "cup", "tree", "toy", and so on (Bloom 2000). Evidently, the concept of a whole physical object is an important one for folk metaphysics to analyze. Thus, it would be a mistake to equate the domain of *conceptual* analysis with the domain of *linguistic* analysis.

I conclude that there is no satisfactory way to promote a public or community-wide conception of concepts to the primary, or central, position in

[4] Terence Horgan and colleagues develop a semantic approach to application intuitions in which semantic competence plays a prominent role (Graham and Horgan 1998; Henderson and Horgan 2001).

the project of conceptual analysis. From an epistemic standpoint, certainly, it is best to focus on the personal psychological conception of concepts as the basic starting point, and view the public conception of concepts as derivative from that one in the indicated fashion.

8. *Are intuition-based beliefs justified a priori?*

Defenders of intuition-driven methodology hold that intuitions provide evidence, or warrant, for classification propositions of interest to philosophers. What kind of warrant is this? The warrant in question is commonly held to be of the a priori variety. Intuition, after all, is a traditional hallmark of rationalism, an oft-mentioned source of a priori warrant. Is this something I am prepared to accept? Isn't my purpose, in this and related papers, to show how the evidence-conferring power of intuitions fits within a naturalistic perspective in epistemology? How can a priori warrant be reconciled with epistemological naturalism?

A first reply is that, in my view, there is no incompatibility between naturalism and a priori warrant. True, many contemporary naturalists, following Quine, wholly reject the a priori. But I see no necessity for this position. My favored kind of epistemological naturalism holds that warrant, or justification, arises from, or supervenes on, psychological processes that are causally responsible for belief (Goldman 1986, 1994). The question, then, is whether there are kinds of psychological processes that merit the label "a priori" and are capable of conferring justification. It seems plausible that there are such processes. The processes of mathematical and logical reasoning are salient candidates for such processes. They are processes of pure ratiocination, which is the hallmark of the a priori. So I see no reason why epistemic naturalism cannot cheerfully countenance a priori warrant (Goldman 1999).[5]

It is an additional question, however, whether arriving at classification intuitions is a species of a priori process, and whether it gives rise to belief that is warranted a priori. This must be examined carefully. We must first distinguish between first- and third-person uses of application intuitions to draw conclusions about concepts. Start with the third-person perspective on application intuitions.

[5] A main theme of naturalistic epistemology is that the project of *epistemology* is not a (purely) a priori project. But it doesn't follow from this that there is no a priori warrant at all.

Concept-analyzing philosophers seek the intuitions of others as well their own. Third-person conceptual investigation can readily be interpreted as a proto-scientific, quasi-experimental enterprise, the aim of which is to reveal the contents of category-representing states. Under this quasi-experimental construal, each act of soliciting and receiving an application judgment from a respondent may be considered a complex experimental procedure. The experimenter presents a subject with two verbal stimuli: a description of an example and an instruction to classify the example as either an instance or a non-instance of a specified concept or predicate. The subject then makes a verbal response to these stimuli, which is taken to express an application intuition. This intuition is taken as a datum—analogous to a meter reading—for use in testing hypotheses about the content of the concept in the subject's head. From the point of view of the experimenter, the philosopher engaged in conceptual analysis directed at another person, the evidence is distinctly observational, and hence empirical. The warrant he acquires for any belief about the subject's concept is empirical warrant.

What about first-person cases, where a philosopher consults his own intuitions? This is where a priori warrant looks most promising. In consulting one's own intuition, one makes no observation, at least no perceptual observation. Does this suffice to establish that any warrant based on the intuition is a priori warrant? No. Although the inference from non-observational warrant to a priori warrant is often made, I think it's a mistake. Some sources of warrant are neither perceptual nor a priori. One example is introspection; a second is memory. Introspection-based warrant about one's current mental states is not a priori warrant; and memory-based warrant about episodes in one's past is not a priori warrant. Since some sources of warrant are neither perceptual nor a priori, application intuition might be another such source.

Indeed, the process of generating classification intuitions has more in common with memory retrieval than with purely intellectual thought or ratiocination, the core of the a priori. The generation of classification intuitions involves the accessing of a cognitive structure that somehow encodes a representation of a category. Of the various sources mentioned above, this most resembles memory, which is the accessing of a cognitive structure that somehow encodes a representation of a past episode. Thus, although I am perfectly willing to allow that application intuitions confer warrant, I don't agree that the type of warrant they confer is a priori warrant.

9. *Kornblith's critique of "détente"*

In this final section I briefly respond to Hilary Kornblith's critique of my approach as presented in earlier papers. Kornblith (this volume) argues that the "détente" I offer between methodological naturalism and the method of appeals to intuition just won't work. There are three strands to his argument. The first concerns the question of whose concepts philosophers should analyze, and whether intuitions should be uncontaminated by theory (i.e., as Kornblith interprets it, whether the preferred concepts should be pre-theoretical). The second concerns the question of whether there is any point to the project of studying commonsense epistemic concepts as a precursor to the study of scientific epistemology. I have defended the value of studying commonsense concepts, as a first stage of philosophizing. Kornblith disputes its importance. Third, Kornblith claims that standard philosophical analysis is committed to the thesis that concepts are mentally represented as necessary and sufficient conditions, the so-called "classical" view of concepts. This view, Kornblith tells us, has been refuted by empirical psychology. So here is a sharp conflict between empirical findings and traditional philosophical methodology. How can I hope to achieve a détente between empirical psychology and traditional philosophical methodology when the two approaches conflict so sharply?

On the first point, Kornblith argues against the view that we should study just the intuitions and concepts of the folk. On the contrary, he urges, the theory-informed intuitions of thoughtful philosophers should count for more than the intuitions of the folk (who have given no systematic thought to a philosophical topic). Furthermore, in contrast to the methodological precept that urges suspicion of theory-contaminated intuitions, Kornblith says that theory-informedness is a good thing.

The problem with this argument is that two entirely different relationships are being conflated between theories and concepts (or theories and intuitions). A theory can be related to a concept either by being embedded *in* the concept or by being a theory *of* the concept. A theory *of* a concept says that the concept has such-and-such content. A theory embedded *in* a concept isn't about the concept at all; it's about some other set of phenomena. The intuitional methodology I preach only says that one should avoid intuitions that are influenced by a theory *of* the target concept. Influence by such a theory can prevent the target from issuing a "normal" response to an example, a response that expresses the real content of the concept.

The methodologist's desire to avoid theory-contaminated intuitions should not be confused with a desire to avoid intuitions concerning theory-embedded concepts. There is nothing undesirable about theory-embedded concepts. I part company with Kornblith when he suggests that theory-embedded concepts are *superior* to theory-free concepts, because there are all sorts of theories. A concept that embeds a bad theory is of dubious worth. So I don't share Kornblith's preference for consulting philosophers' intuitions simply because their concepts embed theories more than folk concepts do. The crucial point, however, is the distinction between a methodological stricture against theory-contaminated intuitions and a possible stricture against theory-embedded concepts. I endorse only the former.

Kornblith's second criticism takes issue with my endorsing the study of folk epistemic concepts as a helpful precursor to the study of scientific epistemology. This endorsement was predicated on the idea that we must first identify the features of folk epistemology in order to figure out how it might be transcended by scientific epistemology, while ensuring that the latter project is continuous with the former. Here is an illustration of what I had in mind. Examining folk epistemic concepts should reveal how truth (true belief) is a primary basis of epistemic evaluation and epistemic achievement. This is indicated, for example, by the truth-condition on knowledge and the reliability desideratum associated with justifiedness. When moving from folk epistemology to scientific epistemology, we should retain the concern with truth-related properties of methods and practices. We should try to make them more reliable than our existing practices. If we never studied folk epistemic concepts, or studied them without proper understanding, this desideratum might elude us. It has indeed eluded postmodernists and (many) sociologists of science, who spurn the activity of conceptual analysis applied to concepts like knowledge or justification. They preach a kind of reformed or purified epistemic regime that ignores truth altogether. This radical and unfortunate detour from traditional epistemological concerns could be avoided by not abandoning folk epistemic notions and not neglecting the important features they highlight, such as truth.

Kornblith's third criticism is that a serious respect for the findings of cognitive science is incompatible with traditional conceptual analysis. I cannot advocate both, as I appear to do. Traditional analysis assumes that concepts are represented in terms of necessary and sufficient conditions, whereas cognitive science tells us that concepts take quite a different form from this classical one. Kornblith urges us to heed the teaching of cognitive science and abandon traditional conceptual analysis.

I deny that traditional analysis is committed to the thesis that concepts (in the psychological sense) are mentally represented by features that are individually necessary and jointly sufficient. In fact, in two previous papers (Goldman 1992; Goldman and Pust 1998: 193–194) I have specifically recommended the exemplar-based approach that Kornblith also calls to our attention. The method of consulting intuitions about cases places no constraint on the psychological format of concept representations. *Any*_hypothesis about concept representations that correctly predicts "observed" classification intuitions is tenable and welcome. Intuition-driven methodology imposes no requirement that hypotheses must posit a classical format for concept representation. True, in formulating the content of a concept representation, philosophers have customarily adopted the format of necessary and sufficient conditions, but I see nothing essential about that practice. For example, a recursive format could be adopted instead, using base clauses, recursive clauses, and a closure clause. In any case, exemplar based data-structures, paired with a set of similarity operations, might well yield classification judgments that can be captured in terms of necessary and sufficient conditions. (The conditions might involve a rather tedious set of disjunctions of conjunctions.) So the necessary-and-sufficient-conditions format for expressing a concept's content is neutral with respect to the psychological "syntax" by means of which the concept is psychologically represented (and processed).

Finally, I disagree with Kornblith's claim that commitment to a necessary and sufficient condition style of analysis biases philosophers toward unrealistically elegant or "pretty" analyses and toward dismissal of intuitions that shouldn't be dismissed. He criticizes philosophers, for example, for trying to explain away data that seem to show that knowledge can be false, by appeal to examples like "Most of what the experts know turns out not to be true". Admittedly, epistemologists commonly seek an alternative explanation of such intuitively acceptable utterances, an explanation that explains away the implication of false knowledge. But I see nothing wrong with this. It is plausible to explain such cases by saying that our speech often describes direct or indirect discourse, or propositions that are objects of propositional attitudes, while omitting overt quotation marks or attitudinal operators. In the present case, the utterance probably means something like this: "Most of what so-called experts credit themselves with knowing, or are credited by others with knowing, turns out to be false". Here's another case (due to Richard Feldman, 2003: 13) of a (true) sentence that apparently implies the existence of false knowledge. You are reading a mystery story in which all the clues, until the last chapter, point toward the butler. Only at the end do you learn that the accountant did it. After finishing the book you say, "I knew all along that the butler did it, but then it turned out that he didn't". Pursuing the explanatory scheme suggested above, one might paraphrase the sentence as follows: "All along I was prepared to say, 'I know that the butler did it', but then it turned out that he didn't". This is a good explanation of how the sentence is understood, and it doesn't imply the falsity of what was known. This simple explanation of an apparent departure from the rule that knowledge is true looks like a perfectly good maneuver. It offers a general principle of language use that has considerable appeal and makes sense of the indicated utterances. It doesn't look implausibly ad hoc, and certainly not driven by an *unreasonable* commitment to necessary-and-sufficient-conditions-style analyses.

So, to summarize this last section, Kornblith hasn't given us good reason to think that taking cognitive science seriously forces us to abandon the intuitional methodology of conceptual analysis, at least if this methodology is understood in the liberal way I have sketched.

REFERENCES

Alston, W. (1993). *The Reliability of Sense Perception*, Cornell University Press, Ithaca, NY.

Bealer, G. (1998). 'Intuition and the Autonomy of Philosophy', *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, (eds.) M. DePaul and W. Ramsey, Rowman & Littlefield, Lanham, MD, 201–240.

Bloom, P. (2000). *How Children Learn the Meanings of Words*, MIT Press, Cambridge, MA.

Cummins, R. (1998). 'Reflection on Reflective Equilibrium', *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, (eds.) M. DePaul and W. Ramsey, Rowman & Littlefield, Lanham, MD, 113–128.

Dehaene, S. (1997). *The Number Sense*, Oxford University Press, New York.

Devitt, M. (1994). 'The Methodology of Naturalistic Semantics', *Journal of Philosophy* 91, 545–572.

Feldman, R. (2003). *Epistemology*, Prentice Hall, Upper Saddle River, NJ.

Gettier, E. (1963). 'Is Justified True Belief Knowledge?', *Analysis* 23, 121–123.

Goldman, A. (1986). *Epistemology and Cognition*, Harvard University Press, Cambridge, MA.

— (1992). 'Epistemic Folkways and Scientific Epistemology', *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, MIT Press, Cambridge, MA, 155–175.

— (1994). 'Naturalistic Epistemology and Reliabilism', *Midwest Studies in Philosophy* 19 *Philosophical Naturalism*, (eds.) P. French, T. Uehling and H. Wettstein, University of Notre Dame Press, Notre Dame, IN, 301–320.

— (1999) 'A Priori Warrant and Naturalistic Epistemology', *Philosophical Perspectives* 13 *Epistemology*, (ed.) J. Tomberlin, Blackwell, Boston.

— (2005). 'Kornblith's Naturalistic Epistemology', *Philosophy and Phenomenological Research* 71, 403–410.

Goldman, A. and Pust, J. (1998). 'Philosophical Theory and Intuitional Evidence', *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, (eds.) M. DePaul and W. Ramsey, Rowman & Littlefield, Lanham, MD, 179–197.

Graham, G. and Horgan, T. (1998). 'Southern Fundamentalism and the End of Philosophy', *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, (eds.) M. DePaul and W. Ramsey, Rowman & Littlefield, Lanham, MD, 271–292.

Henderson, D. and Horgan, T. (2001). 'The A Priori Isn't All That It Is Cracked Up to Be, but It Is Something', *Philosophical Topics* 29 (1–2), 219–250.

Hintikka, J. (1999). 'The Emperor's New Intuitions', *Journal of Philosophy* 96(3), 127–147.

Jackson, F. (1998). *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Clarendon Press, Oxford.

Kornblith, H. (2002). *Knowledge and Its Place in Nature*, Oxford University Press, Oxford.

— (2005). 'Replies', *Philosophy and Phenomenological Research* 71, 427–441.

— (this volume). 'Naturalism and Intuitions', *Grazer Philosophische Studien* 74, 27–49.

Kripke, S. (1980). *Naming and Necessity*, Harvard University Press, Cambridge, MA.

Locke, J. (1694). *Essay Concerning Human Understanding*, 2nd edition, *Personal Identity*, (ed.) J. Perry (1975), University of California Press, Berkeley, 33–52.

Putnam, H. (1975). 'The Meaning of "Meaning"', *Mind, Language and Reality*, Cambridge University Press, New York, 215–271.

Sosa, E. (1998). 'Minimal Intuition', *Rethinking Intuition: The Psychology of Intu-*

ition and Its Role in Philosophical Inquiry, (eds.) M. DePaul and W. Ramsey, Rowman & Littlefield, Lanham, MD, 257–269.

Weinberg, J., Nichols, S. and Stephen S. (2001). 'Normativity and Epistemic Intuitions', *Philosophical Topics* 29 (1–2), 429–460.

Grazer Philosophische Studien
74 (2007), 27–49.

NATURALISM AND INTUITIONS

Hilary KORNBLITH
University of Massachusetts, Amherst

Summary

This paper examines the relationship between methodological naturalism and the standard practice within philosophy of constructing theories on the basis of our intuitions about imaginary cases, especially in the work of Alvin Goldman. It is argued that current work in cognitive science presents serious problems for Goldman's approach.

In an important series of papers (Goldman 1992a; 1992b; 2005; this volume; Goldman and Pust 1998), Alvin Goldman has sought to defend the philosophical practice of constructing theories on the basis of appeals to intuition. This philosophical method is certainly not lacking for adherents; indeed, George Bealer (1993) refers to it as the "standard justificatory procedure" in philosophy. More than this, the practice of appealing to intuitions is not some unexamined aspect of philosophical practice: quite the contrary, this particular feature of philosophical methodology has recently been the focus of a good deal of attention[1], with quite a number of philosophers offering detailed defenses for their preferred method of theory construction. Many of these philosophers, however, are deeply opposed to naturalism, and their defense of the method of appeals to intuition is a crucial component of their anti-naturalistic worldview. Goldman stands out in this company as a committed methodological naturalist, someone who has regularly argued for the relevance of empirical work to philosophical theory construction. And this, of course, raises a question about the relationship between naturalism and the method of appealing to intuitions: to what extent are these truly compatible?

[1] Especially since a conference at the University of Notre Dame in April, 1996 on this topic, organized by Michael DePaul and William Ramsey. See DePaul and Ramsey 1998. See also Jackson 1998; BonJour 1998; and Pust 2001.

The term "naturalism" is used in a very wide range of different ways, and I will not attempt to legislate here that the term be used in some particular, and inevitably controversial, manner. Instead, I want to examine Goldman's view on philosophical method in some detail, and I will argue that there are important tensions to be found there[2], both internal to the view itself, and also between Goldman's view about philosophical method and his actual philosophical practice. Since these tensions all turn on recognizably naturalistic features of Goldman's larger philosophical commitments, the problems raised here should be of broad philosophical concern. What is at issue is how we ought to proceed in the project of theory construction in philosophy.

I have a positive proposal to make as well, and I will offer a sketch of a naturalistic approach to philosophical method which avoids the problems I see in Goldman's view[3]. It should be pointed out at the beginning, however, that Goldman's approach to methodological issues fits far better than my own with a great deal of recent philosophical practice. For that very reason, the case for my preferred view depends quite strongly on ruling out the possibility of the kind of detente Goldman offers between methodological naturalism and the method of appeals to intuition. If I am right, we are all faced with a starker set of choices among philosophical methods than may have initially seemed to be the case.

I.

Appeals to intuition play a foundational role in a good deal of philosophical theory construction. Consider, for example, one of Gettier's famous cases (Gettier 1963). A hypothetical case is described in which an individual arrives at a belief that p or q on the basis of extremely good evidence that p, but no evidence at all about q. It is stipulated that the belief that p or q is true, but not for the reason the individual in question believes. As it turns out, p is false, although q is true. The individual's belief that p or q is thus a justified, true belief. Nevertheless, as almost everyone who hears this case allows, we have the very strong intuition that this individual does not know that p or q. We thus seem to have a case in which there is

[2] Philip Kitcher suggested a related tension in Goldman's work in his (1992): 69, note 46.

[3] I have articulated and defended this proposal at greater length in my (2002).

justified, true belief, but not knowledge. The claim that knowledge just is justified true belief is thus shown to be false.

How does intuition play a role in defeating the proposed JTB analysis of knowledge? According to Goldman (Goldman and Pust 2002: 182), there are two separate steps here. First, the intuition that the individual in the hypothetical case does not know is taken as evidence that such an individual would not have knowledge. Intuitions are thus treated as evidence for the truth of their contents. Second, the claim that such an individual lacks knowledge is then brought to bear on the proposed analysis of knowledge as justified, true belief. In this case, the intuition is used to defeat a proposed philosophical analysis, but this is not essential to the case. Intuitions may also be used to support proposed analyses, by showing that they square with our intuitions better than available alternatives.

Let us call this method of theory construction in philosophy, following George Bealer, the "standard justificatory procedure".[4] Goldman wishes to endorse a qualified version of this procedure. In particular, Goldman wishes to allow that this kind of appeal to intuitions is legitimate, and that intuitions are, under the right conditions, evidence for the truth of their contents. The standard justificatory procedure, as Goldman sees it, rightly plays a central role in philosophical theory construction. At the same time, the standard justificatory procedure, according to Goldman, cannot be the whole story about philosophical analysis. While appeals to intuition illustrate the armchair character of one aspect of philosophical theorizing, Goldman also insists (see esp. Goldman 1992a) that there is an important place in philosophical analysis for explicit appeals to the results of empirical experimentation. This is where the methodological naturalism comes in.

Goldman, like many others, sees the target of philosophical analysis as a concept. When Gettier proposed his famous examples, he helped us to understand the concept of knowledge, and, in general, philosophical analysis is understood as an analysis of our concepts. Philosophy of mind thus studies our mental concepts; epistemology studies such concepts as those of knowledge and justification; moral philosophy studies such concepts as those of the good and the right. And on Goldman's view, concepts are psychologically real. Thus, my concept of knowledge plays a role in explaining, for example, how it is that I recognize cases of knowledge, how

[4] Bealer insists that the deliverances of intuition are a priori justified, but I do not build that in to my description here. It is a requirement which Goldman explicitly rejects.

it is that I am able to entertain thoughts about knowledge, and how it is that I am able to make certain inferences from claims about what various individuals know.

Since concepts are psychologically real, on Goldman's view, they are susceptible to empirical study, and there is a great deal of experimental work going on (see, e.g., Smith and Medin 1981; Keil 1989; Murphy 2002) which is designed precisely to allow for a deeper understanding of the various features of our concepts. But if someone believes, as Goldman does, that concepts are psychologically real, and also that there is a well established tradition in experimental psychology which studies them, then what room is left for the armchair methods of philosophers, methods designed to illuminate the very same target?

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.

This is, of course, itself a substantive empirical theory about the nature of our concepts and their causal relations, and it is susceptible to empirical investigation. In addition, even if this view is correct, it does not suggest that the armchair methods philosophers use are, by themselves, sufficient to understand the contours of our concepts. Even if the connection between concepts and intuitions is highly reliable, experimental investigation may reveal details of our concepts which intuition does not; it may even serve to correct mistakes which a reliance on intuition alone would produce. Nevertheless, Goldman is quite optimistic about the results of unaided armchair methods in philosophy. While experimental work must be taken seriously as a source of additional detail and possible correction, on Goldman's view, unaided armchair methods are highly reliable.

Goldman's view thus combines a certain conservatism about philosophical method with a methodological naturalist's commitment to the relevance of empirical experimentation to philosophical theory. Such a combination, it seems, provides the best of both worlds. Traditional philosophical methods have, I believe, on almost everyone's views, provided a good deal of illuminating results. By endorsing the standard justificatory procedure, Goldman has a straightforward explanation of why it is that such methods may lead to valuable theory. At the same time, a powerful case has been made, by Goldman himself, among others, for the relevance of experimental work to traditional philosophical issues. Goldman's approach to philosophical method thus attempts to combine these independently attractive features into a single unified view. I have some doubts, however, about the extent to which such a unification may genuinely be achieved.

II.

The first question I wish to ask about Goldman's approach to philosophical method is this: Whose concepts are we supposed to be characterizing? In much philosophical theorizing, philosophers attempt to discover their own intuitions about cases, and then construct unified theories which successfully capture them, at least to the extent that this is possible. David Lewis (1973: 88) has suggested that this is what he did. On this sort of view, a philosopher attempts to characterize his or her own concepts, without regard for the extent to which these concepts match or nearly match the concepts of other individuals. There may be some real differences across individuals, as far as this method is concerned, but what each person should do, on this view, is characterize the contours of his or her own concepts.

But not all philosophers see things in quite this way. Frank Jackson sees the enterprise of conceptual analysis as one in which we characterize certainly widely shared "folk concepts". On this kind of view, it would seem, a good bit of empirical work would need to be done, involving surveys of large numbers of individuals, something which, on its face, it seems that very few philosophers ever do. Jackson's response to this is straightforward:

> I am sometimes asked—in a tone that suggests that the question is a major objection—why, if conceptual analysis is concerned to elucidate what governs our classificatory practice, don't I advocate doing serious public opinion polls on people's responses to various cases? My answer is that I do—when it is necessary. Everyone who presents the Gettier cases to a class of students is doing their own bit of fieldwork, and we all know the answer they get in the vast majority of cases. (1998: 36f)

Goldman seems to suggest a similar idea. When he talks about the process of eliciting intuitions, he remarks,

> A verbal stimulus in two parts is administered to a subject, containing the description of a hypothetical scenario and a question as to whether a certain predicate is satisfied by the scenario. One then observes the subject's verbal response. Many experiments of this kind are performed involving the same predicate, typically with many subjects, and their responses provide evidence concerning the contents of mental representations associated with this predicate or concept. (2005: 408)

Jonathan Weinberg, Steve Stich and Shaun Nichols have actually carried out large-scale surveys of this sort, and Goldman remarks that, "This can be viewed as a more rigorous variant of what philosophers customarily do, which highlights the proto-scientific character of the traditional procedure" (2005: 408).

Let me note one additional point about the eliciting of intuitions. Goldman, and many others as well, have urged that the intuitions we rely on must involve "spontaneous application of the concept uncontaminated by an intuitor's prior theorizing, if any"[5]. This is, I believe, quite an important point. If the targets of philosophical analysis are either an individual's pre-theoretical concepts, or the widely shared pre-theoretical concepts of the folk, then the vast majority of actual philosophical practice probably does a very bad job of getting at them. When a philosopher examines his or her own intuitions, the subject of this experiment is someone who has engaged in a great deal of theorizing about the subject under study, and there is every reason to believe that background theory will affect the subject's intuitions about cases in ways that reveal, not the pre-theoretical concept, but the successor concept which has been shaped, over a period of years and sometimes decades, by extensive theorizing. If one is genuinely interested in pre-theoretical concepts, one could not find a more biased and inappropriate set of experimental subjects than philosophers who devote their careers to theorizing about the concepts under study. If one wants pre-theoretical intuitions, one may easily find subjects who have not devoted a lifetime to the very sort of theorizing which gets in the way of revealing pre-theoretical intuitions. Thus, if the target of philosophical analysis involves pre-theoretical concepts, we will need to affect a shift in

[5] (2005: 406). Note that the suggestion here is that the intuitions should be entirely innocent of theory, or at least they should be as innocent of theory as possible. This is quite a different suggestion than the unexceptionable point that if the source of a certain intuition I is a belief in a theory T, then the having of the intuition cannot count as evidence for the theory. This latter point does not in any way suggest that we should be interested in intuitions which are prior to all theorizing.

philosophical method. We will need to stop consulting our own intuitions, and consult the folk whose intuitions remain uncontaminated.

What about Jackson's suggestion that questions in class are useful in eliciting folk intuitions? I have my doubts about this. First, this is not done nearly as often as would be appropriate were we serious about the idea of getting intuitions from those uncontaminated by theory. Second, the classroom situation is not at all revealing, I believe, of pre-theoretical intuitions. Questions about intuitions are typically prefaced by a good deal of discussion which is theory-informed, thereby undermining the possibility of canvassing pre-theoretical intuitions. In addition, the public showing of hands, as is common in such situations, is subject to a variety of well-documented biases, such as the anchoring effect. 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. So, once again, we see that the idea of getting at pre-theoretical intuitions does not so much explain the philosophical practice it is designed to defend, as suggest that the practice must be substantially revised. The suggestion that we wish to make use of pre-theoretical intuitions does not sit well with standard philosophical practice.

One possibility here is that those who wish to engage in conceptual analysis should give up the idea that it is pre-theoretical intuitions which really count. The worry about "theory contamination" which Goldman raises is not, I think, unreasonable. Rather, what I mean to be suggesting is that the move to pre-theoretical intuition is, perhaps, the wrong solution to the problem. Thus, consider the worry raised by some philosophers of science that observation may be contaminated by theory. It is not at all difficult to produce examples of cases in which observers with deep theoretical commitments make mistakes in their observations precisely because of the way in which their background views infect their observations. Now one response to this problem would be to insulate observation from theory entirely, or at least as much as possible; philosophers who favor this approach see theory-neutral observation as an ideal. But this ideal has come in for a great deal of criticism, not only because theory-neutral observation is arguably impossible, but, more importantly, because it is not even an ideal to which we should aspire. When accurate theory is allowed to play a role in guiding and shaping observation, the theory-mediated observations which result are far more telling than those which are shielded, to the extent possible, from influence by any theory. Thus, the worry about "theory contamination" here is not solved by any attempt to eliminate the role of theory in influencing observation. Theory mediation is not automatically a source of contamination. Instead of seeking to eliminate the role of theory in affecting observation, we need to be aware that observations in science—and scientific method generally—are theory-mediated throughout. We need to make sure, to the extent we can, that the theories which influence our observations are accurate, and thus attention needs to be focused on the independent assessment of the theories which influence our observations.

A similar approach might be taken to the problem of theory contamination for intuitions. It is not influence by background theory which is the problem, on this view. A problem arises only when the background theory which influences our intuitions is mistaken. Intuitions uninformed by any theory—or only minimally informed by theories common to the folk—would be no more useful here than observations performed by investigators wholly ignorant of relevant background theory in science. We do not go out of our way, in the sciences, to have observations made by individuals so ignorant of relevant theory that their corpus of beliefs contain no theories at all which might threaten to affect their observations. By the same token, one might think that, in philosophical theorizing, consulting the intuitions of the folk, who have given no serious thought to the phenomena of knowledge, justification, the good, the right, or whatever subject happens to be at issue, not only shields the resulting intuitions from the potential bad effects of a mistaken theory, but it also assures that the positive effects of accurate background theory cannot play a role. Those who have devoted a lifetime to thinking about knowledge and justification, for example, are certainly capable of making mistakes, and their theory-mediated judgments about these matters are certainly not infallible. But this hardly suggests that we should, instead, prefer the intuitions, uninformed by any real understanding, of the ignorant. The suggestion that we should attempt to capture pre-theoretical intuition, however, seems to privilege the intuitions of the ignorant and the naive over those of responsible and well-informed investigators. I cannot see why this would be a better idea in philosophy than it is in science.[6]

[6] It is worth noting here that some discussions of the role of intuition in philosophy make a great deal of the role of intuition in mathematics and logic. One thing about intuitions in those fields is perfectly clear: it is the intuitions of highly trained and thoughtful investigators which are useful, not the pre-theoretical intuitions of the uninformed folk. To the extent that the analogy with intuition in mathematics and logic is apposite, the suggestion that it is pre-theoretical

If we give up the idea that the intuitions we consult should be "pre-theoretical", and instead endorse a practice of appealing to the intuitions of well-trained and thoughtful investigators, then the result looks far more like standard philosophical practice, in which philosophers appeal to their own intuitions and attempt also to account for the intuitions of their colleagues, than the highly modified philosophical practice which would result from taking the idea of "pre-theoretical intuition" seriously. Those, such as Goldman and Jackson, who defend not only the standard justificatory procedure, but also the ideal of appealing to pre-theoretical intuition, cannot, I believe, have it both ways; typical philosophical practice does not involve appealing to pre-theoretical intuitions. So much the better, I believe, for typical philosophical practice.[7]

If we allow, however, that the theory-informed intuitions of thoughtful philosophers count for more than the intuitions of the folk who have given no systematic thought to a particular philosophical topic at all, then the suggestion that the target of philosophical analysis is a mental representation, i.e., a concept, begins to lose the plausibility which it initially had. Theory-informed judgments in science may be more telling than the judgments of the uninformed because accurate background theory leads to more accurate theory-informed judgment. The uninformed observer and the sophisticated scientist are each trying to capture an independently existing phenomenon, and accurate background theory aids in that task. Experts are better observers than the uninitiated. If the situation of philosophical theory construction is analogous, however, as I believe it is, then we should see philosophers as attempting to characterize, not their concepts, let alone the concepts of the folk, but certain extra-mental phenomena, such as knowledge, justification, the good, the right, and so on. The intuitions of philosophers are better in getting at these phenomena than the intuitions of the folk because philosophers have thought long and hard about the phenomena, and their concepts, if all is working as it should, come closer to accurately characterizing the phenomena under study than those of the naive. So on this view the target of philosophical analysis is not anyone's concept at all. Instead, it is the category which the concept is a concept of. Philosophers, on this view, seek to understand

intuition in philosophy which should play a role in theory construction is thus undermined.

[7] Which is not to say that I endorse typical philosophical practice, as I make clear below. Rather, I think that appealing to the intuitions of responsible and knowledgeable investigators is a better practice than appealing to the intuitions of the folk; but I believe there are still better practices available, and it is these which we should adopt.

knowledge, for example, not anyone's concept of it. Some concepts will more nearly get the phenomenon right than others, so not all concepts are equally good guides to the phenomenon, but our concepts are merely the vehicles by way of which we might come to understand the extra-mental phenomena, not the targets of our researches themselves.

More than this, once we start to see the target of philosophical understanding as an extra-mental phenomenon, we will need to modify the standard philosophical procedure. At best, the standard procedure might help us elucidate the contours of our concepts, and, under the current proposal, it is the concepts of philosophers which are most relevant, rather than the concepts of the folk. But since our ultimate target is the extra-mental phenomenon on this view, we would do better to study those extra-mental phenomena directly rather than to study our own, admittedly theory-informed, concepts. The investigator who is interested in aluminum will learn little about his target by studying folk concepts of aluminum. He will learn more by studying the concepts of sophisticated chemists. But he will do better still to look at aluminum itself, for only then will he be able to get beyond the limitations and potential misunderstandings which are embodied in the concepts of even the experts. Once we allow that some concepts are better than others, and that the concepts of those who are well-informed are likely to be superior to those of the naive and the ignorant, it thus becomes quite difficult to defend the view that it is the concepts themselves which are the targets of philosophical study.

Some, including Goldman, will think that my example, involving the concept of aluminum, is not at all apposite. Aluminum, it will be said, is a natural kind, but the same cannot be said of the subject matter of philosophical investigations: knowledge, for example, is not a natural kind.[8] Now as a matter of fact, I have argued at some length elsewhere (Kornblith 2002) that knowledge *is* a natural kind, and I believe that the same is true of many other topics of philosophical investigation. But I need not insist on that here. It really doesn't matter, for present purposes, whether knowledge and other targets of philosophical analysis are natural kinds. Suppose that the category of knowledge is somehow socially constructed. On this view, the standards for knowledge are not somehow given by the world. They are not something we might discover, existing independently of our preferences and choices; instead, they are imposed on the world by

[8] Thus, in his (1992b: 144), Goldman writes, "Whatever one thinks about justice or consciousness as possible natural kinds, it is dubious that knowledge or justificational status are natural kinds".

human beings. Even if this is true, and even if the topics of philosophical interest typically correspond to such socially constructed kinds, it remains true that the concepts of the folk, and the concepts of philosophers as well, need not accurately characterize these socially constructed categories. Just as any individual's concept of aluminum may contain substantial errors or omissions, any individual's concept of a semiconductor, or Chippendale furniture, or of socially constructed categories generally, may contain substantial errors or omissions. And the same is true of the folk concepts of these categories. So the gap between concept and category does not disappear simply because we have moved from natural kind concepts to socially constructed ones (see Kornblith forthcoming a; forthcoming b). And once we recognize that our concepts, whether the concepts at issue are those of the folk or of theoreticians, may fail to characterize the categories they are concepts of, the philosophical interest of our concepts thereby wanes. Our concepts change over time, as the emphasis on avoiding concepts "contaminated" by theory reveals; no doubt the pre-scientific folk concept of knowledge is quite different from the current folk concept, for example.[9] A focus on our concepts may be of anthropological interest, but it is hard to see why philosophers, qua philosophers, would be interested in our concepts of knowledge, or justification, or justice, rather than knowledge, justification and justice themselves. This point is entirely independent of the contention that knowledge (and other philosophical subjects of interest) constitutes a natural kind.

III.

In 'Epistemic Folkways and Scientific Epistemology', Goldman lays out two different tasks for an epistemological theory: the first, the task of conceptual analysis, is "to elucidate commonsense epistemic concepts and principles: concepts like knowledge, justification, and rationality, and principles associated with these concepts" (1992b: 155); the second, the project of scientific epistemology, "is the formulation of a more adequate, sound, or systematic set of epistemic norms, in some way(s) transcending

[9] Moreover, as this example reveals, the idea that we might get at concepts which are genuinely pre-theoretical by consulting contemporary folk is deeply suspect. The concepts of contemporary folk are just as theory-mediated as the concepts of contemporary philosophers. Moreover, the problem would clearly not be solved by looking to the folk concepts of early hominids.

our naive epistemic repertoire" (1992b: 156). The concepts which emerge from a proper scientific psychology may, perhaps, include some which look like refinements or developments of our folk concepts, but they will by no means be limited to concepts of this sort. I am, of course, completely convinced of the importance of a scientific epistemology; I am unconvinced, however, about the need for an understanding of our epistemic folkways. I hope the remarks I have made thus far serve to cast some doubt on the project of clarifying the nature of our folk epistemic concepts. Goldman, however, provides two different considerations in defense of this project, and I want to examine them directly.

On Goldman's view, it is not just that the task of elucidating our epistemic folkways is a worthwhile project within epistemology, on a par with the project of constructing a scientific epistemology. Instead, understanding our epistemic folkways is a necessary precursor to a proper scientific epistemology. The project of providing a conceptual analysis of our folk concepts, on this view, plays a foundational role. Thus, Goldman remarks,

> Whatever else epistemology might proceed to do, it should at least have its roots in the concepts and practices of the folk. If these roots are utterly rejected and abandoned, by what rights would the new discipline call itself "epistemology" at all? It may well be desirable to reform or transcend our epistemic folkways … But it is essential to preserve continuity; and continuity can only be recognized if we have a satisfactory characterization of our epistemic folkways. Actually, even if one rejects the plea for continuity, a description of our epistemic folkways is in order. How would one know what needs to be transcended, in the absence of such a description? So a first mission of epistemology is to describe or characterize our folkways. (1992b: 155f)

Goldman thus offers two reasons for the importance, and, indeed, the primacy of an analysis of folk epistemic concepts: (1) the concepts of a scientific epistemology can only be shown to be worthy of the name if it can be shown that they preserve the relevant sort of continuity with our folk concepts, and this requires that we have a characterization of what those folk concepts are; and (2) if, as the proponents of a scientific epistemology insist, we are likely to transcend many of the features of our folk epistemology, we will have to know, first, what those features are, in order to transcend them. Let me examine each of these justifications in turn.

The idea that a scientific epistemology must, in some sense, be continuous with our folk epistemology if it is to be worthy of being called "epistemology" at all is certainly a plausible one, but there are weak and strong versions of this idea, depending on just how much continuity is required. Consider, for example, the relationship between alchemy and chemistry. Chemistry did not, of course, come about in an instant, and the transition from alchemy to chemistry, on any reasonable account, shows a continuous development. Some alchemical concepts were retained or modified as chemistry emerged; others were entirely abandoned. If the relationship between our folk epistemology and a scientific epistemology is thought of as comparable to the relationship between alchemy and chemistry, then a similar sort of continuity should be expected. And I certainly have no objection to this at all.

But Goldman seems to have something far stronger in mind. On Goldman's view, contemporary epistemologists need to begin their work by engaging in a conceptual analysis of our folk epistemic concepts. And if the relationship between folk epistemology and scientific epistemology is modeled on the relationship between alchemy and chemistry, it is hard to see why we should demand any such thing. Chemists do not need to begin their investigations by providing detailed analyses of folk chemical concepts. A scientific chemistry certainly aims, among other things, to transcend the confines of our folk chemical concepts, but this does not mean that effort needs to be spent on elucidating those folk categories before the real work of chemical theorizing can begin. Note, in particular, that the philosophical analysis of folk epistemic concepts is a highly contested affair. As many in this tradition have painted it, the project of providing an analysis of knowledge can be seen to undergo continuous development from Plato's *Theaetetus* to Descartes' *Meditations*, through Gettier, right up to the present day. There are hotly contested issues about the analysis of the folk concept of knowledge, and if we need to engage with these issues before we can even think about how to go about transcending our folk concepts, then we are likely to be diverted from the project of a scientific epistemology for quite some time. But chemistry did not have to wait for a detailed conceptual analysis of folk chemical concepts before it could begin. Surely the project of providing a detailed analysis of the folk concepts of matter, or substance, or fire, air, earth and water, would have been a distraction from the project of developing a proper scientific chemistry. The continuity between folk concepts and scientific ones takes care of itself as inquiry proceeds; we needn't devote special effort to elucidating our folk concepts in order to achieve it.

A proper scientific epistemology will need to be continuous with our folk epistemology if it is to be worthy of the name, but the kind of continuity required does not serve to motivate the project of providing an analysis of our folk epistemic concepts. If the relationship between the concepts and principles inherent in our folk epistemic ideas and the concepts and principles of a proper scientific epistemology is anything like the relationship between other folk concepts and principles and their scientific successors, then the cause of understanding in epistemology will only be advanced by direct pursuit of the scientific project, leaving the task of providing an understanding of pre-scientific concepts and principles to the historians of the field.

IV.

Thus far I have argued that the project of analyzing our folk concepts does not plausibly aid in achieving the kind of philosophical understanding we seek. Those, such as Goldman, who make the case for a scientific epistemology, and a scientific understanding of various other philosophical topics, should not, I have argued, value the project of providing an analysis of our folk concepts. But this is not my only concern about Goldman's philosophical method. As noted earlier, Goldman sees the standard justificatory procedure in philosophy as a proto-scientific endeavor, continuous with the experimental investigation of our mental representations. As Goldman sees it, the results of conceptual analysis as done from the armchair are unlikely to be modified very much when experimental results are consulted. Experimental work is relevant here, and some modification and correction of armchair theorizing is inevitable. But Goldman's project is not to undermine the standard justificatory procedure in philosophy; it is, instead, to legitimate it. So the thrust of Goldman's project is ultimately conservative: it is to defend a very largely traditional philosophical methodology by way of non-traditional means.

Now I certainly think that there is a valuable scientific project in coming to understand the nature of mental representation, and a great deal of work has been done in pursuit of such understanding. But, at least as I read this literature, it does not help to justify the traditional armchair methods of philosophers, or anything even roughly like them, as a way of elucidating the contours of our concepts. And if this is right, then even if there is some philosophical benefit to be had in understanding the character of our folk concepts, this should not motivate us to adopt the traditional armchair methods of philosophy.

When philosophers attempt to provide an analysis of folk concepts, they typically try to provide a set of necessary and sufficient conditions in a certain standard form.[10] Knowledge, for example, may be analyzed, on certain views, as justified, true belief meeting some additional, and difficult to specify, condition. It is taken for granted that the form of a proper analysis is just some such set of individually necessary and jointly sufficient conditions. The idea that our concepts are mentally represented in this form is what psychologists refer to as the Classical View of concepts. Since the early to mid-1970's, it has become increasingly clear that the Classical View is not correct. Some of the most important problems of the Classical View are due to the discovery of typicality effects. I will give one illustration of these.

Consider the folk concept of being a bird. Subjects can classify various animals as either falling under the concept or not, and they have strong intuitions about whether various individuals are or are not birds. One might, on the basis of this, attempt to figure out the set of necessary and sufficient conditions for being a bird—perhaps having wings, a beak, and two legs, or some such thing—and hypothesize that most individuals mentally represent the category of birds by way of such a list of necessary and sufficient conditions for kind membership. If we were to do this, we would be engaging in the very sort of philosophical project which we have been calling the standard justificatory procedure.

But in addition to recognizing individuals as either falling under the concept or not, subjects also tend to regard certain individuals as more typical of the category than others. Thus, for example, robins and sparrows are regarded as highly typical of the category of birds; ostriches and penguins are regarded as highly atypical. There is, interestingly, a great deal of intersubjective agreement on these judgments of typicality. More than this, reaction time experiments show that individuals which are highly typical of a given category are more quickly recognized as members of that category than atypical individuals, even when relevant traits are equally salient. Thus, for example, subjects will more quickly judge that a given robin is a bird than that an ostrich is, even when the wings, beak, and so

[10] The qualification, "in a certain standard form", is crucial here. Any account of concepts will give necessary and sufficient conditions for application of the concept, including prototype and exemplar accounts. What is at issue is not whether there should be necessary and sufficient conditions provided, but the form that those conditions take.

on of the ostrich are as salient as the same features of the robin. If concepts are represented by a list of necessary and sufficient conditions, then judgments about category membership would be arrived at by checking for each of the necessary features. When the features of two individuals, however typical or atypical of the kind, are equally salient then, judgments of kind membership should be arrived at in equal amounts of time. But they are not.

If concepts are represented, however, by a typical case together with a number of dimensions along which category members might vary, as well as permissible degrees of variation for each of these dimensions, then one would expect that typical cases would be recognized more quickly then atypical cases. In fact, speed of recognition of kind membership turns out to be proportional to degree of typicality, something completely at variance with the Classical View of concepts.[11] The manner in which concepts are represented will do more than dictate the class of individuals which are members of the kind; it will also have implications for the manner and speed at which information about the category is processed. There are a number of different views available today about the way in which concepts are represented. But the importance of typicality effects, and a number of other results, have led to the demise of the Classical View. As Gregory Murphy notes in his recent book, "To a considerable degree, it has simply ceased to be a serious contender in the psychology of concepts" (Murphy 2002: 38).

This leads to serious trouble for the traditional philosophical project of conceptual analysis (as noted, for example, by Stich 1998: 104), and especially for Goldman's attempt to reconcile that project with a realistic account of concepts. Our best current theories tell us that concepts are not represented in the standard form of a set of necessary and sufficient conditions in the way that analytic philosophers have typically offered their conceptual analyses. Such analyses almost certainly distort the shape of the concepts we actually have. Moreover, if it is folk concepts which we are actually interested in, as Goldman suggests, then the practice of philosophers to consult their own intuitions and the intuitions of their colleagues is deeply mistaken, because these concepts are shaped by a lifetime of theorizing and thereby offer up a target quite different from the concepts of the folk. Even in those cases where philosophers do consult

[11] Eleanore Rosch did some of the earliest and most influential work in developing this idea. For an early review of the literature, see Smith and Medin 1981.

with the folk, such as in Jackson's remark involving classroom examples, it seems that what philosophers do with this data is entirely different from the project of elucidating the concepts of the individuals consulted. Thus, consider the familiar case of the introductory student who insists that knowledge is compatible with false belief: "Most of what the experts know turns out not to be true", such a student might say. When confronted with this kind of remark, philosophers will typically try to explain it away, offering various reasons for dismissing the intuition as unrevealing of the subject's concept. Now whatever one's theoretical project, individual bits of data will sometimes need to be explained away rather than accounted for, but the reasons for which philosophers dismiss this kind of data is, I believe, revealing. Comments like this are typically dismissed because they fail to fit in to a unified account of the contours of the subject's concept, on the assumption that the form of such an account will consist of a set of necessary and sufficient conditions. The closest one can get, it may be argued, requires dismissing intuitions such as this one as merely aberrant, the product of some sort of interfering factors. But if concepts are not represented in the form of a set of necessary and sufficient conditions, but instead, as some have suggested, in a manner which is based on a set of exemplars (see Smith and Medin, 1981: Ch. 7; Murphy 2002: Ch. 4), then concepts need not have the unity which serves as the basis for dismissing the intuition. Subjects such as this may simply have certain exemplars of knowledge which involve false belief. Perhaps such subjects have concepts of knowledge which are not terribly unified. But in this respect, this is probably true of most folk concepts. The illusion of unity is a product of the false presupposition that concepts are represented by way of a set of necessary and sufficient conditions. Once we give up the idea that concepts are represented in this traditional form, any attempt to characterize folk concepts will need to show far more respect for many of the folk intuitions which philosophers typically explain away, at least when they stop to note their existence at all. Folk concepts are not a pretty thing, but if we are to take seriously the idea that they are the targets of philosophical analysis, we will need to stop cleaning them up and start presenting them for what they are, in all their splendid disunity.

Of course, there is another possible source for the idea that the target of philosophical analysis will prove to be more unified than I am now suggesting. If it is not folk concepts which we are trying to characterize, but the categories they are concepts of, and if these categories do in fact exhibit a good deal of theoretical unity, in the way that, for example, natural kind categories do, then the accounts we offer should reflect the unity to be found in the phenomena, even if our concepts of the phenomena are not nicely unified. But Goldman rejects this view. The elegant and well-behaved philosophical analyses which Goldman himself offers do not look like accurate characterizations of the rather rude and irregular features of our folk concepts. Goldman's view about the target of philosophical analysis does not fit well, I believe, with his own philosophical practice.

One further point about actual philosophical practice deserves discussion here. Those who engage in conceptual analysis typically construct imaginary examples to test our intuitions, and the character of these imaginary examples varies dramatically. Some of them are humdrum everyday cases, cases which, even if they haven't actually occurred, very well might. Such cases are not merely imaginable, they are nomologically possible, and they do not inhabit possible worlds very distant from our own. But some examples which philosophers appeal to involve counterfactual cases describing events and possibilities very remote from the features of our world. There are cases involving Swampmen, creatures who magically coalesce from the particles of a swamp in such a manner that they are particle-for-particle duplicates of real individuals; there are appeals to cases of various deities, whose mental states and processes do not obey laws even remotely similar to those found in the actual world. Philosophical practice varies, of course. Some philosophers freely appeal to cases of this sort; others tend to avoid them. Goldman endorses the more permissive approach, and he takes it as a virtue of his account of philosophical method that it allows for this practice.

> [My] reconstruction of standard practice makes straightforward sense of our willingness to consider hypothetical examples such as the deity and the Swampman examples. Any imaginable case can shed useful light on our concept of knowledge, so long as the concept can be applied to the case and can generate intuitive responses, thereby indicating something relevant about that concept's contours. This highly permissive approach … often strikes scientists as quite odd (because their own projects are of a different nature … (2005: 406)

Now it is entirely straightforward why such cases would be helpful if our concepts took the form of a set of necessary and sufficient conditions. Since the individually necessary and jointly sufficient conditions need only be checked against the description of the imaginary case, any imaginary case is useful, so long as it has enough detail to allow for each of the necessary conditions to be checked against it. It is no complaint against such an example that it describes a situation very remote from the actual world.

Things are different, however, if our concepts take the form of a typical case together with a set of dimensions of similarity, or, instead, if our concepts are founded in a set of exemplars. Judgments about kind membership are ultimately founded on assessments of similarity, if our concepts take either of these forms. Assessments of similarity, however, are extremely sensitive to changes in background information. Cases involving nomological backgrounds utterly different from either typical cases or any of the exemplars actually encountered are unlikely to provide us with reliable information about the contours of such concepts. Just as counterfactual conditionals about worlds very distant from our own often prove difficult to assess, similarity judgments involving cases from such worlds are an uncertain guide to our concepts. Intuitions about such cases are likely to be especially prone to interference from priming effects, features of emphasis and salience, as well as peculiarities of an individual's background views. They will be far less revealing of the contours of our concepts than cases closer to home.

Goldman's permissive views about the kinds of cases which ought to be used in coming to understand the features of our concepts make philosophical method, as Goldman points out, quite different from the methods found in the sciences. But this highly permissive approach is only licensed by a commitment to the Classical View of concepts, and there is very good reason to believe that the Classical View is simply mistaken.

If we genuinely wish to understand the contours of our concepts, as Goldman suggests, then the practice of philosophers does not provide us with a useful method to achieve this understanding. The standard justificatory procedure in philosophy, with its permissive appeals to intuition, presupposes a mistaken view about the logical form of our concepts. More than this, this mistake has very broad implications. Standard philosophical method does not even approximate a reliable method of understanding our concepts. If it is understanding of our concepts which we seek, then we will need to abandon this method entirely, and adopt the experimental techniques of the cognitive sciences.

V.

Let me pause to take stock. I have argued that our concepts are not plausibly viewed as the target of philosophical understanding. Our concepts are a reflection of our understanding of various extra-mental phenomena. Folk concepts exhibit the imperfect understanding of the folk; their concepts, and their intuitions, are often a product of ignorance and error. But this does not suggest that we should, instead, be interested in the concepts of those who are better informed. Rather, it is the extra-mental phenomena themselves which are the real targets of philosophical analysis: knowledge, justification, the good, the right, and so on, not anyone's concepts of these things. I have further argued that the attempt to view concepts as the target of philosophical analysis fits badly with our best current theories of concepts. The standard philosophical procedure cannot be redeemed by viewing it as an attempt to provide an understanding of our mental representations instead of the phenomena which they are representations of. Thus, whatever one's view of the target of philosophical analysis, the standard justificatory procedure will require extensive modification.

What then do I propose instead? Let me very briefly sketch an account of philosophical method which I have elaborated upon in detail elsewhere (Kornblith 2002). I will discuss a single example—the case of knowledge—but I believe that the example generalizes quite broadly.

As I see it, epistemologists are interested in knowledge, rather than our concept of knowledge. There is a certain phenomenon which we seek to understand, and we should approach it directly rather than by way of attempting to understand anyone's view of it. As it turns out, the phenomenon of knowledge has been an object of study for some time, not just by philosophers, but also by cognitive ethologists. Ethologists attribute mental states to various animals in order to explain their behavior, but they also talk about animals knowing various things. An examination of the motivation for this talk reveals that it is more than just a convenient manner of speaking. Talk of knowledge plays an important explanatory role in ethological theory.

The behavior of individual animals is, as I've said, explained by their beliefs and desires. When we seek to explain the presence of various cognitive capacities in a species, however, knowledge enters the picture. Cognitive capacities are adaptations. The environment makes certain informational demands on a species, and cognitive capacities answer to those demands. They are reliable capacities for the uptake of information about the environment, and if we want to know why some individual has a certain cognitive capacity, we will need to advert to the evolutionary explanation for the presence of the capacity in the species. Such capacities are responsive to selection pressures because having true beliefs allows animals to act on their biological needs in ways that are likely to fulfill them. Animals ignorant of their surroundings are less likely to satisfy their needs, and thus, less likely to survive. There is thus a good deal of evolutionary pressure on a species in favor of reliable cognitive capacities for the production of true belief. The category of beliefs which are a product of such capacities, and which in fact fulfill the purpose for which these capacities were selected—reliably produced beliefs which are also true—is thus important to ethologists. As I see it, the explanatory importance of the category and its theoretical unity provide reasons for viewing it as a natural kind.

In this case, and many others as well, I would argue, we may view the targets of philosophical interest as natural kinds, susceptible to straightforward empirical inquiry. The standard approach in philosophy of beginning with individual cases is no different in kind than selecting clear-cut cases of a natural kind for further investigation. What we seek to understand is what it is that the obvious instances of the kind have in common. Figuring out what the various instances of the kind do have in common, however, should not be seen as simply a matter of consulting our intuitions about cases. The features which members of a kind share, and in virtue of which they are rightly seen as instances of a single kind, may be ones which are not immediately apparent. Semantically competent speakers of a language which contains terms referring to the kind need not have any knowledge of these features. Just as it was a discovery that gold has atomic number 79, we may, through empirical research, discover the essential features of the kinds which are of longstanding philosophical interest.

If what I've said here is even roughly correct about some of the targets of philosophical interest, then the standard philosophical method will need to be revised. As I've argued here, however, whatever one's views about the targets of philosophical interest, the standard method will require substantial revision. The kind of conservative approach to philosophical methodology which Goldman and many others favor will not, I have argued, hold up to scrutiny. The only remaining possibilities, then, involve highly non-trivial changes in philosophical practice.[12]

[12] I am grateful to Alvin Goldman for discussion of these issues over a period of many years. I have also had helpful discussions on these issues with Stephen Stich and Jonathan Vogel, as well as with the audience in Erfurt.

REFERENCES

Bealer, G. (1993). 'The Incoherence of Empiricism', reprinted in *Naturalism: A Critical Appraisal*, (eds.) S. Wagner and R. Warner, University of Notre Dame Press, 163–196.

BonJour, L. (1998). *In Defense of Pure Reason*, Cambridge University Press.

Depaul, M. and Ramsey, W. (eds.) (1998). *Rethinking Intuition: The Psychology of Intuition and Its Role in Philosophical Inquiry*, University of Notre Dame Press.

Gettier, E. (1963). 'Is Justified True Belief Knowledge?', *Analysis* 23, 121–123.

Goldman, A. (1992a). 'Psychology and Philosophical Analysis', reprinted in *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, MIT Press, 143–153.

— (1992b). 'Epistemic Folkways and Scientific Epistemology', *Liaisons*, 155–175.

— (this volume). 'Philosophical Intuitions: Their Target, Their Source, and Their Epistemic Status', *Grazer Philosophische Studien* 74, 1–25.

— (2005). 'Kornblith's Naturalistic Epistemology', *Philosophy and Phenomenological Research* 71, 403–410.

— (manuscript). 'Philosophical Intuitions: Their Source and Epistemic Status', presented at the Troiseme Cycle Romand de Philosophie, Fribourg, Switzerland.

Goldman, A. and Pust, J. (2002). 'Philosophical Theory and Intuitional Evidence', reprinted in A. Goldman: *Pathways to Knowledge: Public and Private*, Oxford University Press, 73–94.

Jackson, F. (1998). *From Metaphysics to Ethics: In Defence of Conceptual Analysis*, Oxford University Press.

Keil, F. C. (1989). *Concepts, Kinds and Cognitive Development*, MIT Press.

Kitcher, P. (1992). 'The Naturalists Return', *Philosophical Review* 101, 53–114.

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

Kornblith, H. (forthcoming a). 'How to Refer to Artifacts', *Creations of the Mind: Essays on Artifacts and their Representation*, (eds.) E. Margolis and S. Laurence, Oxford University Press.

— (forthcoming b). 'Appeals to Intuition and the Ambitions of Epistemology', *Epistemology Futures*, (ed.) S. Hetherington, Oxford University Press.

Lewis, D. (1973). *Counterfactuals*, Harvard University Press.

Murphy, G. L. (2002). *The Big Book of Concepts*, MIT Press.

Pust, J. (2001). 'Against Explanationist Skepticism Regarding Philosophical Intuitions', *Philosophical Studies* 106, 227–258.

Smith, E. E. and Medin, D. (1981). *Categories and Concepts*, Harvard University Press.

Stich, S. (1998). ‘Reflective Equilibrium and Cognitive Diversity’, in Depaul and Ramsey, 95–112.

Grazer Philosophische Studien
74 (2007), 51–67.

INTUITIONS: THEIR NATURE AND EPISTEMIC EFFICACY

Ernest SOSA
Brown University and Rutgers University

Summary
This paper presents an account of intuitions, and a defense of their epistemic efficacy in general, and more specifically in philosophy, followed by replies in response to various objections.

I.

What is intuition? What accounts for its probative force? Traditionally, intuition is understood on a perceptual model. It is through the mind's eye that we gain insight. In perception one's eyes may come to rest on an object seen in good light. A sensory experience then mediates between object seen and perceptual belief. However, nothing like sensory experience seems to mediate analogously between facts known intuitively and beliefs through which they are known. Moreover, many truths known intuitively lie outside the causal order, unable to cause experience-like intuitions, even if there were such intuitions. Nor can such truths be tracked, not if tracking requires sensitivity. What are we to make of the claim that if it were not so that 1 + 1 = 2, one would not believe it to be so? Hard to say, but that is what tracking it with "sensitivity" would require.

Even if there are no experience-like intuitions, intuitive seemings remain distinctive conscious states in their own right, without collapsing into beliefs, as is shown by paradoxes like the liar, or the sorites. Each proposition in a paradoxical cluster exerts a powerful intuitive attraction, despite how compelling it also is that they cannot all be true together. Even when one eventually settles on a solution, moreover, the pull of the rejected proposition is not removed but overcome.

What then might intuitions be, if they are to be conscious states with probative force despite being fallible, while distinct from beliefs?

II.

Elsewhere I defend a conception of intuition as a state distinct from and prior to both belief and knowledge.[1] Intuitive belief is based on intuition but goes beyond it, and in turn constitutes intuitive knowledge only if all goes well. In what follows I will presuppose such a conception of intuitions as intellectual seemings of a certain sort, as attractions to assent derived from the sheer understanding of the propositions involved.

Here we focus on *propositional* intuition, which has the following features:

a. It is a conscious state.
b. It has propositional content.
c. It is distinct from belief. One can have an intuition that p without believing that p, as when one resolves a paradox by disbelieving one of its elements, despite the powerful intuitive appeal that it retains.
d. Its content can be false; there can be false intuitions.
e. It does not derive just from perception, introspection, testimony, or inferential reasoning, singly or in combination, not even through the channel of memory.
f. It can serve as a basis for belief, helping thus to provide epistemic justification for the supported belief.

An intuition is hence a representationally contentful conscious state that can serve as a justifying basis for belief while distinct from belief, not derived from certain sources, and possibly false.

Thus:

> S intuits that p if and only if S's attraction to assent to <p> is explained rationally by two things in combination: (a) that S understands it well enough, (b) that <p> is true.

How well does this account of intuition fit the profile above? Quite well on the whole, or so I will argue. On this account, intuition is a conscious state of felt attraction rationally explained through the content's being (a) understood well enough by the subject, and (b) true; and such a conscious state can serve as a justifying rational basis for belief, *ceteris paribus*.

[1] 'Intuitions and Truth', in Greenough and Lynch 2006.

It is instructive to consider *how* we can tell as extensively as we can what someone else believes in given circumstances. Based on our own perception of her situation, and the placement and orientation of her sense organs, we can tell a lot about what she is likely to believe. Thus if I draw your attention to a square plainly visible on a surface and ask you whether it is a triangle, or a circle, or a square, etc., I can easily predict your answer. Moreover, I can explain why you answer as you do, assuming your sincerity, good eyesight, and understanding of the question, by appeal to the sheer truth of the matter, to the fact that the figure *is* a square. If it had been a circle, or a triangle, etc., I would have been able to predict and explain similarly *mutatis mutandis*. But not if it had been a chiliagon!

Somehow I am able to know ahead of time the proper bounds of this sort of explanation. I know that it works with pentagons, and maybe as far as octagons or thereabouts; after that, most people need to stop and count. Moreover, there is an immense amount that each of us knows through such direct perception of the facts, and that each of us knows that nearly everyone else knows similarly when appropriately situated. A lot of what we know in a given situation we know others share with us when thus situated. What accounts for this? Do we depend on some assumption of simplicity? Do we say that others will know the things we know about the perceptible situation so long as they are properly endowed and situated, so that their relevant competence enables them to tell the relevant truths?

III.

There is an important further reason for thinking that rational intuitions must be true. Consider a belief derived as a conclusion from certain reasoning. If that reasoning is fallacious, we would presumably deny that it provides justification for its conclusion. Perhaps it can still help give subjective justification for believing that conclusion, if for example the subject has taken care and knows himself to be normally good at such reasoning. Nevertheless, in some more objective sense we would deny that he is really justified in so concluding. One cannot attain epistemic justification through fallacious reasoning.

When we work our way back through the reasoning, suppose we find a fallacious step of affirming the consequent. Something of the following form at that point seemed intuitively right to our subject: that, necessarily, if $q \& (p \rightarrow q)$ then p. What of his justification for so assenting, for taking that step of immediate inference? Is he there justified through immediate intuitive appeal? Well, he may be *subjectively* justified, in which case he would presumably retain *such* justification for assenting to the eventual conclusion of that reasoning. As we have seen, nevertheless, the reasoner must also *lack* some more objective justification for assenting to his conclusion. It is this more objective justification that he also must lack for assenting to that fallacious step (or to its corresponding conditional). Fallacious intuition cannot secure that stronger sort of justification. Only *true* logical intuitions could provide such justification, or so we may reasonably conclude, since anything less would leave the same problem.

Although that is a reasonable conclusion, it is not unavoidable. Here is an alternative: namely, that the defective intuition involved in affirming the consequent is fallacious because the subject proceeds in some avoidably defective way. It's a *fault* in the subject that he has that intuition, an individual flaw, or defect. It is not just an inaccuracy. By contrast, the false intuitions involved in deep paradoxes are not so clearly faults, individual flaws, or defects. For example, it may be that they derive from our basic make-up, shared among humans generally, a make-up that serves us well in an environment such as ours on the surface of our planet.

Compare the subject who sees lines in a Mueller-Lyer pattern for the first time, with no prior knowledge of that sort of illusion, while *in fact* the lines *are* incongruent; or compare the subject who takes a wall to be red while it is indeed red, though it would have looked red even had it been white, because the light is red, which she has no reason to suspect. In these cases we take her to be epistemically justified in believing as she does. Yet, the explanation for why she so believes is not that the content of her belief is true, even though the content *is* in fact true. In each case she might very easily have believed as she does, even if her belief had been false, so long as the situation had remained misleading in the ways specified.

Here in any case is for us the more important point. Let us now suppose the belief to be *false*, while the situation remains misleading in the ways specified, with the misleading arrows at the ends of the lines, or with the misleading light shining on the surface. In that case, the subject's perceptual belief would still be epistemically justified, yet this can no longer be explained through the truth of its content. At one level the explanation here appeals to the fact that the subject still undergoes a sensory experience sharing its content with the belief based on it as its reason. And this supposedly gives him epistemic justification. Can this be just a brute fact? Preferably we should try to explain it.

Consider the defective appearance involved in one's attraction to think the lines incongruent, or in one's attraction to think the wall red. Does the subject proceed in a remediably, avoidably defective way? Well, yes, in a way; yet remediation is not really an option on her available basis for belief. Is it a *fault* in the subject that she is then pulled so powerfully to think the wall red? Is it an individual flaw, or defect? Surely not. As with the false intuitions involved in deep paradoxes, appearances thus induced are not faults, nor individual flaws, or defects. They seem to derive rather from our basic make-up, shared among humans generally, a make-up that serves us well on the surface of our planet.

That is not to say that the subject will be justified in taking her appearances at face value. Moreover, even if she *is* so justified, this can change with further experience: for example, once she measures the lines, or examines the light. What spontaneously draws her assent, moreover, once she has that further experience, can also change, at least in strength. Getting wise to the effect of the angles at the ends of the lines might reduce the spontaneous attraction to assent. Nevertheless, one's visual module will not be denied, but will have its way in and through a stubborn visual appearance of incongruence, even if now only by bucking firm belief to the contrary.

The important question for us is how to understand the epistemic justification of belief that the lines are incongruent, or belief that the wall is red, in the circumstances respectively specified. Might we appeal to the ability or competence or faculty or cognitive virtue manifest in the formation of that belief? We explain the formation of true, apt beliefs in whole fields of true propositions by appeal to such abilities seated in a subject. But these abilities need not be thought infallible, which they would have to be if we defined them as abilities always to discern the true from the false in those fields. Our abilities might be highly reliable without being infallible. Our perceptual abilities seem clearly fallible, given that circumstances can so often be misleading while about this we have no clue, and cannot be expected to have a clue. This remains true even when the ability depends, for its proper, successful operation, not just on pertinent circumstances, but also on the subject's mental shape at the time.

A belief formed by such an ability might still be considered epistemically justified, when the subject is in bad circumstances or in bad shape. This suggests that our evaluation of the act (the assent, or, in another context, the choice) may indirectly involve an evaluation of the agent, and of her character so manifest, abstracting from the low quality of her internal or external situation. In the case of vision, and more specifically of congruence vision, or color vision, it seems possible, indeed likely, that the virtues manifest in the relevant inclinations to believe, and in the assents and beliefs themselves, should be viewed not as infallible, but as fallibly dependent for their success on conditions of the external situation and on *further* features of the subject's mind. Suppose that, through no fault attributable to the subject, things go wrong like that, the effect being a false belief. Even so, the belief might be epistemically justified, in the following sense: that it is well-formed, rationally well-motivated, deriving as it does from the exercise of a virtue; that is to say, from a feature of the subject's mind that guides him reliably well in normal conditions.[2]

Applying this to the case of intuition, we have a way to distinguish the blunder of the thinker who affirms the consequent from the non-blunder of the subject with a paradox-enmeshed belief. The fallacious step, we can now say, is *unjustified*, in that it depends on the shape of the subject's mind at the time in a way that was easily enough remediable, and not epistemically excusable. It depended on the subject's inattentiveness, since no normal thinker who is sufficiently attentive will commit that kind of blunder. By contrast, paradoxes reveal something much more like deep-seated perceptual illusions to which the visual system itself is inevitably prone. Thus, perhaps the same rational virtue that accounts for our prowess at telling the true from the false in the field of simple enough a priori propositions, is inevitably prone to errors that manifest themselves in deep paradoxes. If so, if our intuitions in the paradoxical cluster do all derive from that same rational faculty, then they can all enjoy *prima facie* epistemic justification, even if one of them at least must be false. Familiar facts about deep paradoxes render that account plausible. Intuitions constitutive of such a paradox can hardly be explained through inattentive blundering, or through any other of the factors to which we appeal in explaining why we reason fallaciously.

Finally, the paradoxes seem in this way more persuasive indictments of our reasoning competence than any of the sorts of mistakes uncovered more recently by psychologists. After all, these latter sorts of mistakes are rather like individual blunders due to inattention. These we are able to correct and avoid through better concentration, which distinguishes them

[2] However, it might be best to distinguish *human* justification from the superhuman justification of a Ramanujan. Human justification would derive from human faculties, shared by normal people.

from how we go astray in traditional paradoxes. Even *if* the difference is a matter of degree, moreover, it is a big difference nonetheless.

So we see how paradoxes challenge the infallibility of intuition. And the infallibility of introspection seems no less questionable. I may look within and think there are nine dots in my visual field where in actual fact there are only eight. This would seem still to count as introspection despite the fact that it goes astray.[3]

Eventually we want to understand rational intuition and intuitive justification more generally. Are they fallible? Descartes did not think so. And there are reasons to agree, some already noted in our discussion of fallacious reasoning. On the other side, consider the paradoxes. These plausibly give rise to examples of false justified intuitions. Only certain special intuitions are protected against fallibility. Similarly, we can feel introspectively confident that we host a certain attitude, with *prima facie* justification, even if our intuitive justification is outweighed by evidence from our behavior that shows our confidence to be misplaced.

IV.

Justified introspective and rational intuitions can be false, as can perceptual intellectual seemings and even justified perceptual beliefs. These latter can be explained by appeal to unfavorable external conditions. But how are we to explain the former?

In what way does our fallacious reasoner fall short of justification? In what way might this be different from the paradox-enmeshed intuition?

Rationally well motivated choice and belief involves *reason*, and what is due to its operation, what not. Compare *mis*remembering. How are we to understand this? We might define retentive memory of <p> so that the following is a necessary condition:

> At t, the subject believes <p> and at t′, later than t, the subject believes <p> because she believed <p> at t.

[3] Suppose, for another interesting case, that we focus on the length of the lines in a Mueller-Lyer *image*. I mean the relative length of the lines as they appear in one's visual image. Are they congruent? Are they incongruent? Neither answer seems obvious, yet there plausibly *is* a correct answer. If so, we have here a way in which one might go introspectively wrong. Or take a victim shown a red-hot, smoking iron, soon after which a very cold iron is pressed to his bare back. He screams in apparent pain. But it is not immediately obvious what to say about this case.

If so, then no instance of the following will count as *mis*remembering, because none will count as a case of memory at all (assuming no sort of memory except the retentive is in play):

> At t, S believes <p>, and at t′ (t′ > t), S believes <p′> because she believed <p> at t, where <p′> is very similar to <p>.

Thus, for example, if at t' you believe that your friend's phone number is 352-2792 because at t you believed it to be 352-3792, this will not count as misremembering, since it will not count as remembering at all.

By contrast, common sense allows the fallibility of memory. When our new belief about the phone number differs from the earlier belief by only one of seven digits, this would be considered a case of misremembering.

We can thus see a way to think of rational and introspective intuition so that these are fallible, which suggests the following modified account:

> S rationally intuits that p if and only if S's attraction to assent to <p> is explained by a competence (an epistemic ability or virtue) on the part of S to discriminate the true from the false reliably (enough) in some subfield of modally strong propositional contents that S understands well enough, with no reliance on introspection, perception, memory, testimony, or inference (no further reliance, anyhow, than any required for so much as understanding the given propositional content).
>
> S introspectively intuits that p if and only if S's attraction to assent to <p> is explained by a competence (an epistemic ability or virtue) on the part of S to discriminate the true from the false reliably (enough) in some subfield of self-presenting propositional contents that S understands well enough, with no reliance on ratiocination, perception, memory, testimony, or inference (no further reliance, anyhow, than any required for so much as understanding the given propositional content).[4]

[4] Note that I speak here of propositions, though more properly it should be propositional contents, in order to take proper account of cogito effects, where indexicals and demonstratives must be given their due.

V.

We have arrived at a general account of intuition as a source of epistemic justification, a conceptual specification of what it is, one that plausibly accommodates probative force for such intuitions. What remains to be seen is how far such intuition can take us in philosophy. To offer positive support here incurs the appearance of vicious circularity. What might be done more effectively is to face the objections of those who reject philosophical intuition as useless.

One main objection derives from alleged disagreements in philosophical intuitions, ones due in large measure to cultural or socioeconomic or other situational differences. This sort of objection is particularly important and persuasive, so I devote a separate paper to discussing it in detail.[5]

It is often claimed that the recourse of analytic philosophy to intuitions in the armchair is in the service of "conceptual analysis". But I find this deplorably misleading, for the use of intuitions in analytic philosophy, and in philosophy more generally, should not be tied to conceptual analysis. Consider some main subjects of prominent debate in analytic philosophy: utilitarian versus deontological theories in ethics, for example, or Rawls's theory of justice in social and political philosophy, or the externalism/internalism debate in epistemology, and many others could be cited to the same effect. These are not controversies about the conceptual analysis of some concept. They seem moreover to be disputes about something more objective than just our individual or shared concepts of the relevant phenomena. Yet they have been properly conducted in terms of hypothetical examples, and intuitions about these examples. The objective questions involved are about rightness, or justice, or epistemic justification. Even if these *are* questions about such objective matters, and not just about our concepts, this does not entail that rightness, justice, and epistemic justification *must* be *natural kinds*. Nor need they be socially constructed kinds. Indeed, we can engage in our three philosophical controversies, and regard them as objective, without ever raising the question of the ontological status of the entities involved, if any. Mostly we can conduct our controversies, for example, just in terms of where the *truth* lies with regard to them, leaving aside questions of objectual ontology.

[5] That paper, 'A Defence of Intuitions', will appear in Bishop and Murphy 2006; a preliminary version is at http://homepage.mac.com/ernestsosa/Menu2.html. Compare also 'Experimental Philosophy and Philosophical Intuition', forthcoming in a collection on experimental philosophy edited by Shaun Nichols and Joshua Knobe.

Moreover, we surely do and must allow a role for intuition in simple arithmetic and geometry, but *not only* there. Indeed, I ask you to consider how extensively we rely on intuition. I myself believe that intuition is ubiquitous across the vast body of anyone's knowledge. Take any two sufficiently different shapes that you perceive on a surface, say the shapes of any two words. If it's a foreign language, you may not even have a good *recognitional* grasp, a good concept of any of those shapes. Still you may know perfectly well that they are different. And what you know is not just that the *tokens* are different: you also know that anything *so* shaped *would* be differently shaped from anything *thus* shaped (as you demonstrate the two shapes in turn). Or take any shape and any color, or any shape and any sound. And so on, and on, indefinitely. There just seems no sufficient reason for denying ourselves a similar intuitive access to the simple facts involved in our hypothetical philosophical examples. That would seem to be the default position, absent some specific objection. There *have* been objections, of course, such as those pressed by Robert Cummins, those pressed by Stephen Stich (most recently in collaboration with Shaun Nichols and Jonathan Weinberg), and those that derive from Paul Benacerraf and Hartry Field. But these all have convincing answers, or so I have argued elsewhere.

Objections and Replies[6]

Objection 1
From the inside, blind prejudice and rational intuition do not feel any different (on the view proposed above). Phenomenologically they seem the same, which gives rise to this worry:

> *How do we know that there is any rational source of this inclination at all (given that they could stem from blind prejudice) unless we have done cognitive research on the relevant processes?*

It may be replied that the case is analogous to introspection. We rely on the assumption that there is a reliable process even if we lack any introspective or other knowledge of it, of its reliability. But it is not so clear that these cases are analogous. There is at least this difference. Introspective judgment is guided by the specific evidence constituted by the conscious state involved, whereas

[6] The following dialectic owes much to a long email exchange with Thomas Grundmann soon after the conference at Erfurt on philosophical methodology.

rational intuition is not guided in any such way.

Suppose, moreover, there are defeaters to your prima facie rational intuition in a particular case. How can you know that these defeaters should be taken seriously? From the first person perspective you feel the strong inclination towards judging that p and, therefore, you judge that p. Suppose that you then have an opposing inclination. As far as you know, however, this inclination might be either the output of a rational source or the result of blind prejudice. As long as you do not know the source of this inclination you do not know whether or not you should suspend judgment. Justified acceptance seems thus to require a fairly direct access to the relevant sources or processes that are operative. According to the externalist, however, one need not know that the source in question is reliable.

My first worry about the position proposed is then this: that it fails to make allowance for a phenomenological difference between rational intuition and prejudice, one that seems required for epistemological purposes.

Reply 1

But won't any of our sources—perception, memory, inference, as well as introspection—have misleading counterparts? Isn't there ostensible perception, memory, etc., that is ostensible only, not real? What makes it *merely* ostensible is presumably that, as far as one can tell introspectively, there is all the appearance that the source is real: it appears just like perception, or just like memory, etc. What then should we require before the source can have its epistemic efficacy? Do we require that the subject must have an ability to tell introspectively in every possible circumstance whether or not the ostensible source is real and reliable, or is merely ostensible and misleading? If we require such *prior* knowledge independently of the operation of that very source, we face the problem of noncircular calibration, and also a problem of vicious regress (as we must tell with priority that the source of belief B is source S1, and we must tell with priority that the source of *this* meta-belief is source S2, and with priority that the source of this meta-meta-belief is source S3, etc. So it seems out of the question to make any such priority requirement in general for our basic sources and their operation in particular instances. Why then pick on intuition for any such special requirement?

Objection 2

Be that as it may, there is a further, more intuitive, worry. Even if we concede to externalism that we need no reflective access to the sources of belief, there *seems a phenomenological difference between prejudice and rational intuition. If I "see" the truth of "Everything is (necessarily) identical to itself" this kind of self-evidence is introspectively different from the kind of pure inclination that a racist may have when he is inclined to believe that his people are superior to the others. In the first case there seems to be more than the naked inclination to judge: there seems to be an additional evidential state.*

Reply 2

Yes, *rational* intuition does have this special phenomenologically distinguishable feature, and my view can accommodate this, through how its distinctive contents relate to modally strong propositions. So, one can appeal to something phenomenologically distinguishable, involving the nature of the propositional content of what one is attracted to accept through mere understanding; rational intuition might involve the modally strong character of what one intuits. (Here I assume that if one properly understands the propositions involved, one will be attracted to assent not only to $2 + 2 = 4$, but also to Necessarily, $2 + 2 = 4$.)

We might also distinguish a subset of contingent intuitions as "animal" intuitions concerning the physical world around us and its physical features, something like basic principles of folk physics, and basic commitments concerning the taking of sensory experience at face value; and also basic intuitive commitments concerning folk psychology, concerning how to "read" faces and gestures in certain ways, for example, commitments that come with our animal endowment (and are not dependent on the specifics of a particular culture's enculturation).

We might even consider the possibility of good culturally dependent intuitions. Certain moral basic commitments might derive from culturally dependent intuition that is correct and epistemically effective. However, if one has doubts about moral realism, and the applicability of the basic epistemic framework to such normative subject matter, then one can stop short with animal intuitions (of the folk-physics or folk-psychology sorts), and one might opt to be even more restrictive, stopping with rational intuitions (concerning simple modally strong subject matter).

Objection 3

There is an important difference between, on the one hand, the individuation of the relevant belief-producing process or method and, on the other hand, the epistemic value of this process or method (dependent on its reliability). In my view, the belief-producing process is individuated internally whereas its *epistemic value depends on external factors, such as its truth-ratio. So even if we cannot and need not tell real perception from merely ostensible perception, we nevertheless have introspective access to the perceptual character of the relevant processes (or at least to their experiential character, perception being factive). For this reason we are always able to tell what the relevant process is like, though we need not be able to say anything about its reliability.*

In short: I think that externalism about the epistemic value or adequacy of our belief-producing mechanisms should go hand in hand with internal access to the kind of source or mechanism that is in fact operative. As I see it this requirement is not satisfied by your account of rational intuition. But it would be satisfied if we identified rational intuition with a process in which distinctive intellectual seemings, analogous to sensory experiential states, are the input.

You do offer a response to these concerns as follows:

> *Yes, I agree that rational intuition has this special phenomenologically distinguishable feature, and my view can accommodate this, through how its distinctive contents relate to modally strong propositions. So, one can appeal to something phenomenologically distinguishable, involving the nature of the propositional content of what one is attracted to assent to through mere understanding; in the case of rational intuition the nature might involve the modally strong character of what one intuits. (And here I assume that if one properly understands the propositions involved, one will be attracted to assent not only to 2 + 2 = 4, but also to Necessarily, 2 + 2 = 4.)*

However, it seems to me that modally strong propositional content is neither necessary nor sufficient for a state to be one of rational intuition. Such content is not necessary since: (1) beliefs about metaphysically far-fetched possibilities might still be based on rational intuition, and I do not see why rational intuition should not also yield beliefs about what is possible. (2) You suggest that anybody who properly understands a necessary proposition will be attracted to think that it is necessarily true. But now consider children or even the layperson having learned basic arithmetic. I think they understand propositions like "2 + 2 = 4" properly, but they may not possess the concept of neccessity. Do you really think that they do not have a proper understanding of mathematical propositions?

Nor is modally strong propositional content sufficient for a state to be one of rational intuition: It is quite easy to conceive of someone who relies on strong blind prejudice in believing that slaves necessarily have to serve their superiors *(that certain slaves are essentially, i.e. necessarily, subordinate people). I don't see why strong prejudice as well as rational intuition cannot have modally strong propositions as propositional content.*

Reply 3
I agree that one can have the intuition that *Possibly p*, but this is itself a proposition with modally strong status. That is, it is itself either necessarily true or necessarily false. As for the arithmetic intuitions, I am not sure that the layperson, and even children who have fairly basic conceptual sophistication, can be wholly innocent of any conception of necessity. Surely they must have some conception of what is possible and what is not possible. This seems essential to minimal planning, and minimal selection of alternative courses of action. And they must also have some conception of negation. But as soon as they have possibility and negation, they must have necessity.

Yes, it does seem possible, however unlikely, that people will have wild prejudices about *essential* (modally strong) racial superiority. But why not say that this just shows that rational intuition is not infallible? And we knew that already because of the ancient paradoxes, such as the sorites.

Objection 4
Here is a further problem for your conception of modally strong propositions. You say:

> *Yes, one can have the intuition that possibly p, but this is itself a proposition with modally strong status, or so I would contend. That is, it is itself either necessarily true or necessarily false.*

If so, would not sentences like "Water is $H_2 0$" or "Cicero = Tully" express modally strong propositions (in your sense) although they do not contain necessity as part of their content? But now I have a problem with your claim that rational intuition can serve as the source of a given judgment through the modal strength in its propositional content. The above judgments clearly seem to be empirical judgments rather than judgments based on rational intuition, although they express modally strong propositions.

Reply 4
Yes, good point. However, my "propositions" are meant to be, not Russellian propositions, but more like Fregean representational contents, or else

combinations of Russellian propositions along with modes of presentation. In any case, instead of the Russellian [Cicero = Tully], the relevant representational content for me would be something like <The Ciceronian = the Tullian> where the two individual concepts involved in the constitution of this "representational content" are different in such a way that it is not a necessarily true content.

The main point is that my representational contents incorporate modes of presentation, or individual concepts, in such a way that the Russellian proposition [Cicero = Tully] does not count as such a representational content. The mode of presentation associated with "Cicero" is different from that associated with "Tully". Accordingly, my account of rational intuition should be interpreted so that it concerns mode-of-presentation-including "representational contents" and does not concern Russellian propositions (except indirectly).

In line with the above, though the relation is complex, I would distinguish between the following two propositions.

1. Necessarily, if there is a unique Ciceronian, x, and there is a unique Tullian, y, then x = y.
2. If there is a unique Ciceronian, x, and there is a unique Tullian, y, and x = y, then Necessarily x = y.

The former is false, the latter true.

Objection 5

Finally, I would like to press a new and different line of objection. In principle it seems always possible that propositions with the same content are based on very different sources or processes. For example, I may know by introspection that I am currently in pain, but someone else may know the same thing only by observation of my behavior. Or I may know (be justified in believing) that there is a pencil on my desk, either by currently observing it or by remembering that I saw it there a few moments ago. So I am inclined to believe that there are no domains of knowledge (or justified belief) classified by the objects or contents. Rather there are different sources or processes that may overlap in the content of their output-beliefs. I do not see why the same does not hold for the sources of propositions with a strong modal content. From my perspective a blind prejudice yielding a necessity claim is not simply a false rational intuition but rather a different (and unreliable) source of belief.

It also is not obvious that there could not be quite a few prejudices (and empiri-

cally based judgments not based on rational intuition) with strong modal content. Think of a statement like the following: "Whites have always had the power and this couldn't be different". Claims about the nature (=essence) of things or people are necessity-claims. But surely not all of them rely on rational intuition.

Reply 5

This too is an important point. But, first of all, one could of course know from Ramanujan's testimony the truth of a modally strong proposition that one must take on his sayso, since one lacks his ability to intuit the truth of that proposition. Nothing in my account precludes that one can have sources other than intuition for one's knowledge of a modally strong proposition, and even for a modally strong proposition that someone else *does* intuit.

Secondly, I fear you may be holding up intuition to an unfairly different standard by comparison with the other sources. There are many, many different perceptual sources. There is short-term memory and also long-term memory, and other sorts of memory as well. Introspection most probably has quite different more specific forms. So, I was willing to distinguish several different forms of intuition, depending on the contents involved (just as we can distinguish varieties of perceptual sources through the kinds of content involved: shape versus color, for example). So there would be the intuitions of folk physics, of folk psychology, of abstract modally strong propositional contents, etc. Different sub-sources, more specific yet, may also be distinguishable. But it is unclear to me why this sort of thing should pose a fundamentally different sort of problem for intuition than for perception.

Objection 6

If I understand you correctly, on your view intuition is a very wide-ranging source. This source produces beliefs (or inclinations to believe) about very different subject matters, modally strong propositions being only one among others. Your response to me, then, amounts to this: as in the case of perception, sub-sources of intuition can only be distinguished from one another by the kind of content involved. So far there is no difference between e.g. perception and intuition. About this I agree. But here is my problem: I don't think that rational intuition is just a sub-source of intuition conceived of more broadly. My reason is simple. From an everyday point of view all the things you have in mind share a certain feature. We are inclined to believe them spontaneously. But on my view this does not indicate a common source of all these inclinations (as there is in the case of all the perceptual sub-sources). The processes leading to these inclinations are not sufficiently similar to each other to constitute a *homogeneous natural (or objective) kind. The real sources are as different as memory, subliminal perception, wishful thinking, prejudice and rational intuition. So, my objection stands: We should be able to distinguish basic sources by phenomenological epistemic features. On the level of basic sources (in contrast to sub-sources) we must not rely just on the content involved. Rational intuition rather than intuition in the everyday sense is a basic source. So we need an epistemic mark of rational intuition. And this is what I miss within your account.*

Reply 6

I doubt that we have here an important substantive disagreement. Perception too is enormously varied, if we include the full panoply of things we normally say we "see" or "perceive", including matters of shape, number, color, physiognomy, all the way to the aesthetic and the moral, and the evaluative more generally. It seems to me that the really important issue is whether we agree that there is a distinctive sub-source for *rational* intuition (whether interpreted in your more restrictive sense, or in my slightly more liberal one). We can perhaps agree to leave out contingent intuitions generally, whether of the folk-physical or the folk-psychological sort, or of any other sort. And the interesting issue will yet remain as to whether some at least of our philosophical intuitions can figure among the broader class of *rational intuitions*. If we agree on this, we really agree on the most important substantive matter. And you seem willing to grant that we could view intuitions as just attractions to assent drawn by sheer understanding, where the subject matter is modal (either by containing a modal concept, in a restrictive view, or by having a modal status easily enough accessible to reflection, in my more inclusive view).

REFERENCES

Bishop, M. and Murphy, D. (2006). *Stich and His Critics*, Blackwell.

Greenough, P. and Lynch, M. (2006). *Truth and Realism*, Oxford University Press.

Grazer Philosophische Studien
74 (2007), 69–87.

THE NATURE OF RATIONAL INTUITIONS AND A FRESH LOOK AT THE EXPLANATIONIST OBJECTION

Thomas GRUNDMANN
Universität Köln

Summary

In the first part of this paper I will characterize the specific nature of rational intuition. It will be claimed that rational intuition is an evidential state with modal content that has an a priori source. This claim will be defended against several objections. The second part of the paper deals with the so-called explanationist objection against rational intuition as a justifying source. According to the best reading of this objection, intuition cannot justify any judgment since there is no metaphysical explanation of its reliability. It will be argued that in the case of intuition the very demand of such an explanation is based on a category mistake.

In everyday language, we use the term "intuition" to refer to a broad range of phenomena: when a state of affairs strikes us immediately as plausible; when we suddenly have the unmistakable feeling that our judgment about something is correct, although we cannot say what it is based upon—like when we predict the development of weather patterns or of the stock market, or when we suddenly foresee some future event, such as the death of a close friend or the success of our job application; when we have a sudden insight or idea; when we respond to a question automatically by giving a memorized answer; when we know exactly how somebody feels, but can't say how we know. Countless other examples could be added. In all these cases, we make an evaluation or judgment without being aware of any inference, perception or memory upon which the judgment is based. This is the criterion according to which most psychologists define "intuition". Gopnik and Schwitzgebel write, for example: "we will call any judgement an intuitive judgement, or more briefly an intuition, just in case that judgement is not made on the basis of some kind of explicit reasoning process that a person can consciously observe" (Gopnik/Schwitzgebel 1998: 77). It is no surprise that psychologists who endorse this definition do not rate the epistemic value of intuitions all too highly. For one thing, intuitions in the everyday sense are not *source-specific*. A judgment whose source is unknown to us could have any one of a number of sources, e.g. memory, experience, background knowledge, wishful thinking, prejudices, guesswork or subliminal perception. The reliability of these sources varies so greatly that one can hardly call intuition as a whole reliable: there are simply too many irrational factors involved in the production of intuitive judgments. It would be equally false to say that intuitions in the everyday sense are independent of experience, or a priori, since—as I just mentioned—they are often covertly guided by perceptions, experiences or empirical background knowledge.

In the Rationalist tradition the attempt has been made to distinguish a philosophical concept of intuition from the everyday, source-unspecific concept. In this philosophical sense, intuitions are not spontaneously experienced judgments but a specifically Rationalist source of evidence. Intuitions are reasons that are characterized by the clear and distinct appearance of truth, and these reasons arise purely a priori. By and large, we say that understanding certain propositions is sufficient to cause an evidential state or clear and distinct insight into the truth, independently of empirical reasons.

Some typical examples of such propositions would be:

(1) Everything is necessarily identical with itself.
(2) Something cannot have the property of being triangular and not have it.
(3) No object can simultaneously be entirely red and entirely green.
(4) 2 plus 2 necessarily equals 4.
(5) If someone knows a proposition p and knows that p logically implies the proposition q, and thus deduces that q, then she necessarily knows that q.
(6) If someone were in an area full of fake barn facades and, in clear view of the one real barn, made the justified, true judgment “That is a barn there,” then she would have no knowledge of this fact.

In these cases, philosophical intuition is directed at logical or epistemic principles and simple mathematic truths, but also concerns the evaluation of counterfactual instances that—as in (6)—are of fundamental importance as test cases in corroborating or impugning philosophical analyses.

In the typical cases, philosophical intuition refers to modal facts. The proponents of Rationalism also hold that modal intuitions are reliable and can justify the corresponding judgments a priori. Although many Rationalists have also claimed that intuitions are infallible, this additional claim is inessential. Today there are hardly any Rationalists who would hold to the infallibility claim. (See Casullo 2003)

Using George Bealer as a starting point, I am going to begin by fending off a few objections to the philosophical concept of intuition as a source of evidence. Then I will address what is at present the most serious objection to intuitions as evidence, namely the explanationist objection. I will try to show that this objection cannot discredit the evidential value of philosophical intuitions.

I.

According to George Bealer, philosophical intuitions are "intellectual seemings" (Bealer 1998; Bealer 2002). They are not judgments or beliefs, but conscious cognitive states with the same propositional content as the judgments based upon them. They differ from these judgments only in their propositional attitude. That, intellectually, it seems to me that p, is not the same as my belief that p. The case of perception is quite similar. That, sensuously, it seems to me that p, does not mean that I believe that p. In a perceptual illusion, the perceptual experience persists even when I no longer believe in my perception because I know that it does not correspond to the facts. On the other hand, believing something does not suffice to make me perceive it with my senses. For Bealer, sensuous evidence is analogous to intellectual seeming in that it is a conscious psychological state that is independent of judgment and has an intentional content. In both cases there is a kind of evidence that is independent of judgment. In other respects, of course, there is a disanalogy between perception and philosophical intuition. Perception has a non-conceptual intentional content, whereas intellectual seeming is obviously conceptual. And, of course, both have different sources. In perception, it is the sense organs and the information they deliver that are decisive; in intellectual seeming it is only a matter of understanding the relevant proposition and the concepts it contains. After all, philosophical intuition is supposed to be an a priori kind of evidence. So the decisive point for Bealer is that philosophical intuitions are independent evidential states. What is obviously true is that they (in contrast to the psychological concept of intuition) are not judgments or beliefs. This point can be illustrated especially clearly by paradoxes where we find various premises intuitively evident but discover that they yield a contradiction. Although we can no longer believe in all the premises, they nevertheless retain their intuitively evident character even in the absence of the corresponding belief. By the same token, beliefs are not automatically intuitively evident, no matter how firmly they may be held. I have calculated that the sum of 1256 and 798 equals 2054. I am firmly convinced of this, but I do not grasp its truth as intuitively evident in the same way that I grasp the truth of "2 + 2 = 4" as intuitively evident. Thus, beliefs are neither necessary nor sufficient for philosophical intuitions.

It does not automatically follow, however, that intuitions are independent evidential states, i.e. intellectual seemings. This is demonstrated by Ernest Sosa's suggestion that intuition be regarded as an inclination or attraction to the corresponding judgment. (Sosa 1998; Sosa Ms.) Such an inclination could exist even if the judgment is not manifest. But the purely dispositional analysis of intuition is relatively implausible, since intuitions are conscious, whereas dispositions exist whether or not we are conscious of them. That is why Sosa introduces an additional condition: intuitions are inclinations to judge of which we are introspectively aware. (Sosa 1998: 259) Moreover, they are based solely upon an adequate understanding of the proposition. According to Sosa, intuitions are introspectively conscious inclinations to judge (if they are based solely upon an adequate understanding of the proposition). Sosa finds this analysis more plausible than Bealer's account of intellectual seemings because he does not think that this supposedly evidential foundation of the inclination to judge is phenomenologically demonstrable.[1] When we consider a proposition, an attraction to assent arises immediately.

However, in his contribution to this volume Sosa characterizes intuitions by the following three features: (i) intuitions are "distinct conscious states", (ii) intuitions are different from beliefs, (iii) and intuitions are "representationally contentful" states with a strongly modal content, i.e. necessity is part of their content. But then there is no big difference to intellectual seemings any more. Of course, intuitions don't have any sensory content. Furthermore, they are spontaneous results of considering the

[1] Sosa Ms.: 12 "No such state of awareness, beyond the conscious entertaining (of the proposition, T.G.) itself, can be found in intuitive attraction".

content of certain propositions, rather than results of external stimulation. But I don't see any reason why Bealer should deny that.

Bealer, Sosa and BonJour characterize the content of intuitions as necessary. So whenever I grasp something through rational intuition, I grasp that this fact is necessary.[2] This thesis seems too strong to me. A core group of central philosophical intuitions refers to the evaluation of counterfactual situations in which assumptions about necessary and sufficient conditions must prove themselves. We imagine a possible situation and ask whether the application of a given predicate (such as knowledge, justification, personal identity or freedom of the will) to this possible situation seems intuitively evident to us. We have the intuition, for example, that the believer in a Gettier-situation does not have knowledge, although we do not have the intuition that she *necessarily* has no knowledge in this situation. (Pust 2000: 38) Hence, rational intuitions have a modal content, but may not contain necessary facts.[3]

Could one perhaps even go so far as to claim that intuitions have no modal content at all but, rather, can be reduced to judgments about actual cases? Why can't we simply find exotic examples in actual history or in distant parts of the actual world and look at our actual judgments in these cases? (Cf. Williamson 2004) This proposal, I think, underestimates the modal, counterfactual dimension of our intuitions. The method of cases does not only serve to undermine philosophical analyses of necessary and sufficient conditions by counterexamples but also may confirm it. If we only consider actual cases, as far-fetched as they may be, there is no way we can confirm conditions of necessity. Even if in the actual world all cases of justified true beliefs were cases of knowledge, that still could not confirm the thesis that justified true beliefs necessarily amount to knowledge. We could be dealing with a universal but contingent co-variation of two properties. So we have to consider as many possible scenarios as we can imagine and ask whether they are cases of knowledge or not. Hence,

[2] Cf. Bealer 1998: 207 "when we have a rational intuition … it presents itself as necessary: it does not seem to us that things could be otherwise". For similar views, cf. also Bonjour 1998: 106-7, Sosa Ms.: 18.

[3] In his contribution to this volume Sosa argues that even the intuition *possibly p* has a strongly modal status in so far as its truth value is necessary. This may be right. But that's not equivalent to the claim that necessity is part of the intuition's content, at least if this content is constituted by modes of presentation. There are necessarily true rational intuitions that don't present us the facts as being necessary as well as there are necessary truths a posteriori that do not present us the facts as necessary. The latter is admitted by Sosa in his reply to my objection 4 (this volume).

the evaluation of actual cases is not sufficient for our purposes. In fact, it is not necessary either, since the only thing that is of relevance for the philosophical analysis is what we *would* say if the situation *were* such-and-such. If, on the basis of the available data, we happen to have been mistaken about the actual facts—if, in other words, the situation turns out in fact not to have been as we had thought—that would not impugn the relevance of our evaluation for the philosophical analysis in the least. It is sufficient if the case is possible. In order to understand philosophical intuition, then, it is essential to take into account its modal, albeit not in all cases necessary, content.

So let's assume that philosophical intuitions are independent evidential states in the sense of Bealer's intellectual seemings, and that they have at least a modal content. Bealer, BonJour and Sosa go one step further in claiming an a priori, albeit fallible, status for this evidence. That is the step at which naturalists like Hilary Kornblith balk. It may well be, they say, that philosophical intuitions have a certain epistemic value. But, since scientific progress may necessitate their revision, they do not represent an ultimate authority. For naturalists, the provisional epistemic value of intuitions can be explained by the fact that they are dependant upon our earlier empirical theories, although we cannot know that introspectively. Intuitions are, as it were, the platitudes of yesterday's empirical background theories. (Kornblith 1998)

But the phenomenological facts oppose this general suspicion. If intuitions were only the platitudes of yesterday's theories, they would in principle have to be conservative. Our received worldview should suffice to explain them. But, interestingly enough, our intuitions often yield counterexamples to the traditional analysis of a phenomenon. Gettier-like counterexamples to our traditional conception of knowledge demonstrate that especially well, as do the Frankfurt-scenarios, which contradict the traditional view that moral responsibility implies the existence of alternate possibilities to act or decide. In such cases, we have revolutionary intuitions that contradict our traditional worldview. (Williamson 2005: 128)

But intuitions do not have to be revolutionary in order to demonstrate their independence from background knowledge. In many cases, we have intuitions even in areas in which we possess no knowledge whatsoever. (Bealer 1998: 209) The slave Menon is probably the primordial case of intuitive acquisition of mathematical knowledge. This impression becomes all the more firm when one recalls that philosophical intuitions have a modal content and have something to say about remote counterfactual situations and, indeed, all possible worlds. It may be that modal knowledge can to some extent be extrapolated from empirical theories, but it seems unlikely to me that all our modal intuitions could ultimately be empirical, i.e. based on knowledge about the actual world.

Williamson (in Williamson 2005) considers the possibility that our usual (i.e. empirical) ability to make counterfactual judgments is sufficient to account for the essentially counterfactual dimension of some counter-examples. He is thinking here above all of counter-examples that bring the metaphysical possibility of particular cases into play. Williamson suggests understanding the metaphysical necessity of A such that for any proposition p: if p were the case, then A ($\Box A = \text{Def.}:\forall p(p \mathrel{\Box\!\!\rightarrow} A)$). Accordingly, the metaphysical possibility of A would be understood such that there is at least one proposition p such that it is not true that if p were the case, A would be false ($\Diamond A = \text{Def.}:\exists p \neg(p \mathrel{\Box\!\!\rightarrow} \neg A)$). If this is right, Williamson thinks, we can base our judgments about metaphysical modalities epistemologically upon usual counterfactual judgments. But, as I see it, our empirically based competency to make counterfactual judgments does not get us all that far. It just so happens that our background empirical knowledge does not enable us to say whether A would or would not be that case if p described a situation that were very remote from our actual situation. Hence, we could not justify (but only, in the best case, refute) metaphysically necessary propositions in this manner, nor could we refute (but only, in the best case, justify) metaphysical possibilities. So I contest that our empirical ability enables us to recognize Williamson's counterfactual basis for reduction of modal judgments. Moreover, Williamson's analysis commits him to saying that we can only recognize necessary truths inductively through counterfactual evaluation of arbitrarily many scenarios. Phenomenologically, though, it seems that we grasp the necessity of at least some truths directly ("Everything is necessarily identical with itself"). If that is so, then it would not be the justification of counterfactual conditionals but the justification of our judgments of necessity that is epistemically primary. Thus, that it is true for any proposition p that, if they were true, A would be true, we recognize *because* we recognize that A is necessary, not vice versa.

I think that should make it sufficiently clear that philosophical intuitions cannot in general be explained by our empirical background knowledge. If we regard clear cases in which intuitions have a non-empirical origin as paradigmatic, then cognitive science should investigate whether the psychological mechanism that operates in these cases always operates when intuitions arise. This question, like the question about the nature of the fundamental mechanism, cannot be answered purely introspectively; both questions demand careful empirical-scientific study. For the moment I am simply going to assume—although, as I say, this assumption could turn out to be empirically false—that in all cases in which an intuition arises, it arises solely because the proposition in question, along with the concepts it contains, has been understood.

II.

Among contemporary challenges to the idea that intuition is a source of evidence the explanationist objection figures most prominently. This objection has been formulated in various ways by Paul Benacerraf, Hartry Field and Alvin Goldman.[4] The core of the objection can be reconstructed in the following form as argument (A):

(P1) If the truth-makers of judgments based on reasons of type X do not (or cannot) play any causal role in the causal explanation of these reasons of type X, then reasons of type X have no justificatory or epistemizing force.

(P2) Modal facts can play no role in the causal explanation of our philosophical intuitions.

Ergo: Philosophical intuitions have no justificatory or epistemizing force.

This argument appears highly plausible at first glance. (P1) satisfies the criteria we commonly accept. We all assume that perceptual reasons justify our perceptual judgments about the world around us and—under favorable circumstances—even lead to perceptual knowledge. Under favorable circumstances, for example, the tree in front of me causes me to have a visual experience of it. From the explanatory perspective, we can explain our experiences of perceptual evidence causally by appealing to the operations of our cognitive-perceptual apparatus. From a broader perspective, one could say that the distal causes of our perceptual reasons are normally objects that correspond to our perceptual judgments. Perception—the paradigmatic

[4] Cf. Benacerraf 1973; Field 1989; Goldman 1992. For a good reconstruction of the argument, see Pust 2000: Ch. 3.

case of a good epistemic reason—satisfies the condition formulated in the first premise. Consider, in contrast, the case of clairvoyance. We do not attribute justificatory or epistemizing force to clairvoyant reasons, regardless of how accurate a clairvoyant's predictions may be. Since, in this case, we don't know any story that causally relates the reasons to the events predicted, the connection appears to us to be purely accidental. In short: the first premise satisfies quite well the criteria we commonly accept.

(P2) is sometimes justified by saying that modal reality per se, as a region of abstract Platonic entities, is causally impotent, or at least that it can have no causal influence upon psychological facts in the natural world, since the natural world is causally closed and allows no external causal influence. (Benacerraf 1973) This argument, however, is dependent upon strong ontological assumptions about the nature of modal reality and of the natural world. Such assumptions are not necessary, though, to demonstrate the causal impotence of modal facts. This is a point made by David Lewis: causal relations are relations that contain a modal dependency between two events or facts. Usually that relation is expressed by the use of a counterfactual conditional: if the cause had not occurred, then the effect would not have occurred either. Now, if you regard modal facts as candidates for causes and assume that what is necessary is necessarily necessary, and that what is possible is necessarily possible, then you will quickly realize that the antecedent of a counterfactual statement that refers to modal facts can never be true. The modal facts simply could not have been other than they are. Thus, according to Lewis, all counterfactual conditionals are vacuously true. But if they are trivially true, then they cannot express the substantial modal dependence between two facts that would be required for causality. Modal facts are therefore causally impotent, regardless of which ontological analysis happens to be on offer. (Lewis 1986: 111)

I think this point demonstrates that (P2) is beyond challenge. Moreover, the argument appears to be valid. What options remain, then, for someone who wants to defend the epistemic value of philosophical intuitions? On the one hand, an intuition freak could try to show that the argument at the core of the explanationist objection is inconsistent. On the other hand, she could attack (P1) directly. If she chooses this course, she again has two options. She could accept the requirement of an explanatory connection between evidence and the corresponding truth-maker, but deny that this connection has to be of a causal nature.[5] Then, of course, she would have

[5] Bealer, BonJour and Sosa offer an explanation in this sense.

to come up with a non-causal explanation of philosophical intuitions. Her second option would be to show that the requirement of an explanatory connection does not have universal scope but, rather, is restricted to reasons for our beliefs about non-modal facts.

Joel Pust has recently made the claim that the explanationist argument is, in a certain sense, self-defeating. (Cf. Pust 2000: Ch. 4) This kind of self-defeat can be characterized more precisely as epistemic inconsistency. An argument is epistemically inconsistent if either (a) its conclusion is inconsistent with the justification of at least one of the argument's premises or (b) one of the premises contradicts its own justification. The conclusion of an argument that contains an epistemic inconsistency can, of course, be true. But an epistemic inconsistency robs anyone asserting the conclusion of her *entitlement* to that claim.

Let's have a closer look at Pust's two arguments that the explanationist objection contains an epistemic inconsistency. The first argument runs as follows:

(1) The conclusion of the explanationist objection impugns the justificatory force of philosophical intuition.
(2) (P1) can only be justified by philosophical intuition.
Ergo: If the conclusion is true, then (P1) is not epistemically justified. (Pust 2000: Ch. 4.3.)

Pust's second argument runs as follows:

(3) If (P1) is true, then all statements about facts that do not play a role in the causal explanation of our evidential states are devoid of epistemic justification.
(4) The facts referred to in (P1) play no role in the causal explanation of our evidential states.
Ergo: If (P1) is true, then (P1) is not epistemically justified. (Pust 2000: Ch. 4.4.)

Self-defeating arguments always have the drawback that they can, in the best case, show that a certain position is false or, at least, that one would not be justified in espousing it, but never *why* that position is false or unjustified. But Pust's arguments fail for a different reason. First of all, premises (2) and (4) can be challenged. Pust thinks that epistemic principles (like that formulated in (P1)) can only be justified by intuition, or at least that they cannot be justified by experience. Both premises rest on this assumption. A naturalist like Kornblith, however, would probably insist that we could justify a premise like (P1) simply by empirically investigating the nature of paradigmatic cases of justified belief. If this is true, then the reason for (P1) satisfies the condition (P1) formulates and is not dependent on intuitions in the philosophical sense. In other words, Pust's argument presupposes the falsity of naturalism. So, in view of the fact that the explanationist objection is raised by naturalists, Pust's counterargument amounts in the end to begging the question.

But even if Pust's arguments could persuade us that the explanationist objection harbors an epistemic inconsistency, it would not automatically follow that the objection falls apart. Let's look at another, comparable case. Putnam once formulated the following meta-inductive argument against induction: if one looks at the history of the sciences, one observes that every inductively based empirical theory has eventually been falsified and abandoned. If all inductively based empirically theories in the past have turned out to be false, then one can induce that all inductively based theories are false. Thus, induction has no justificatory force. This is an epistemically inconsistent argument. If the conclusion is true, then it is not justified, because the argument itself is based upon induction. I can only show that induction cannot justify by using induction. Still, the argument could yield a sort of reductio ad absurdum of the method. If applying a method to itself reveals that the method is unreliable or epistemically worthless, that undermines the justificatory force of the method. A proponent of the explanationist objection could argue in the same parasitic fashion: if one judges philosophical intuitions according to conditions of justification that they themselves support, they do not measure up. That would suffice to undermine the justificatory force of intuitions. As I see it, Pust's self-defeating argument fails for both reasons.

Let us examine the arguments that speak in favor of (P1). *Argument from the causal conception of knowledge*: Early proponents of the explanationist objection, such as Benacerraf, based (P1) upon a causal conception of knowledge. Since, according to Benacerraf, one can speak of knowledge if and only if a true belief (or the reasons for it) were caused by the truth-maker, there cannot be knowledge of the causally impotent modal world. (Benacerraf 1973) But, as is well known, the causal conception of knowledge failed. A causal relation between a true belief and its truth-maker is neither necessary nor sufficient for knowledge. The successor-theory to the causal conception was reliabilism, according to which a belief is justified if and only if the belief-producing process gives rise to true beliefs in most cases in the actual and in relevant, nearby possible worlds. Perfect reliability is required for knowledge. Reliabilist theories contain no causal condition in the definiens, and they cite only necessary conditions for justification. Hence, a reliabilist can undermine Benacerraf's justification of (P1) by denying the casual condition of knowledge. Moreover, the argument can only, in the best case, formulate a condition for knowledge. It has no relevance to the possibility of justification anyway.

Let's take a look at a second argument for (P1). *Argument from old-fashioned foundationalism*: Many old-fashioned foundationalists claim that we have direct knowledge only of intra-mental states—that is, of our percepts, of episodic memory, and also of our intuitive reasons. To these we have introspective access; and if we want to justify beliefs about the external world, we can only rely on an inference to the best explanation. If this picture of our epistemic situation is correct, then it follows directly that we can have no justified beliefs and no knowledge of regions of the external world that have no causal power with respect to our intra-mental states. But, as it happens, this is not a true picture of our epistemic situation. It just is not true that we first have introspective knowledge of our own mental states, and then base our beliefs about the external world upon this knowledge. Rather, we base these beliefs directly upon sensuous experiences, episodic memories or intuitions. The putative intermediate step through introspection is demonstrable neither by phenomenology nor by cognitive science. Justified beliefs about the external world do not presuppose justified beliefs about intra-mental facts. If this is correct, then the evidential connection between our perceptual, mnemonic or intuitive reasons and our judgments cannot be characterized as an inference to the best explanation. On the contrary, in these cases we take at face value the contents of our reasons. Hence the argument considered for (P1) collapses.

And so we come to a third argument for (P1). *The argument from internalism*: Even if the method in question does not itself contain an inference to the best explanation, one can of course ask whether our perceptual experiences or our intuitions are reliable indicators with respect to the facts referred to in the judgments that we base upon them. If we want to answer this question impartially—that is, independently of the method in question—then we will look for an epistemically non-circular meta-justification of these methods. The obvious choice would be an inference to the best explanation. If we ask ourselves whether the beliefs that we base upon perception, memory or intuition arise in a reliable manner, and if we do not want to rely on perception, memory or intuition in answering this question, then we have to consider the possibility that the facts that correspond to our beliefs would provide the best explanation of our perceptual, mnemonic or intuitive evidence. If one also takes an epistemically non-circular meta-justification to be a condition for justification, then one can derive a version of (P1). The argument goes as follows:

(1) A reason can only justify a belief if its reliability is meta-justified in an epistemically non-circular manner.

(2) Only if an inference to the best explanation can be made from the occurrence of a reason to the truth-maker of the belief based on that reason can the reliability of that reason be justified in an epistemically non-circular manner.

(3) The inference to the best explanation from the occurrence of the reason to the truth-maker of the belief based on that reason is only valid if the truth-maker plays a role in the causal explanation of the occurrence of the reason.

Ergo: If the truth-maker of a belief based on a reason of type X plays no role in the causal explanation of the occurrence of type X reasons, then type X reasons have no justificatory force.[6]

Surprisingly, this argument that Goldman proposes for (P1) is purely internalist.[7] For the reliabilist there is no need to accept that a reason has justificatory force only if its reliability is meta-justified in an epistemically non-circular manner. If, on the other hand, this condition is dropped, the argument for (P1) collapses. But if, from an internalist perspective, one were to accept the argument, it would be equally fatal to all basic cogni-

[6] Alvin Goldman's justification of (P1) takes the same course. He writes: "What evidence is there that our possession of these algorithms (Goldman calls in this passage the intuitive procedures ‚algorithms') is somehow related to mind-independent modal facts? The only evidence that has been adduced is our intuitions … And the mere occurrence of these intuitions does not have much probative force once we recognize that there are competing explanations that make no commitment to extramental modal facts. … Isn't it more likely (or at minimum, equally likely) that one of these competing explanatory stories—one that makes no commitment to extra-mental modal facts—is a better explanation, a more reasonable explanation …, than some explanatory story that makes such commitment? … The question is: are our conceptual dispositions … reliable evidence for objective metaphysical fact? Is inconceivability a reliable indicator of impossibility? Unless there is some story that underwrites this indicator relationship, the epistemic status of the intuitions is problematic" (Goldman 1992: 62–3).

[7] Pust, too, is baffled by Goldman's Internalist argumentative strategy; see Pust 2000: 64, Fn. 12.

tive faculties (including perception and memory). Admittedly, perceptual reasons, unlike intuitive ones, can be explained causally by referring to what makes the beliefs based upon them true. But skeptical objections demonstrate that there are equally good alternative explanations (like the brain-in-a-vat or the evil daemon hypothesis) that get by without referring to the corresponding truth-maker. (Alston 1993) In short, this justification of (P1) would imply a global skepticism about the external world. Thus, it does not yield a specific argument against intuition as an epistemic source.

The fourth argument for (P1) is much more promising. *Argument from metaphysics*: as soon as we discover that the reliability of some given reasons cannot be explained with reference to a causal connection between those reasons and the corresponding truth-maker, the justificatory force of the reasons is neutralized. A so-called defeater has come into play. Harty Field formulates this point as follows: "The challenge … is to provide an account of the mechanisms that explain how our beliefs about these remote … entities (or our grounds for these beliefs) can so well reflect the facts about them. The idea is that if it appears in principle impossible to explain this, then that tends to undermine the belief in … (these) entities, despite whatever (initial) reasons we might have for believing in them" (Field 1989: 26; for a similar interpretation see Casullo 2003: 144–45). There is one important question that Field unfortunately does not answer, namely: why does a defeater arise when we discover that we cannot explain a reliable connection between beliefs or reasons and facts? Answering this question will enable us to make the requirements for an explanation more precise.

First: Someone's prima facie reason for one of her beliefs is defeated either if she discovers something that impugns the truth of this belief (i.e., a rebutting defeater) or if she acquires a reason that impugns the reliability of the prima facie reason (i.e. an undercutting defeater). In the case at hand, one thing is clear—namely, that the discovery of an explanatory gap with respect to the reliability of the reason does not impugn the truth of the belief. So we are not dealing here with a rebutting defeater. The other possibility would be an undercutting defeater. But why should the reliability of, say, perceptions as indicators of extra-mental facts be impugned if that reliability cannot be explained? I see here only one course to take—namely, to say that the reliability of a reason is *not a brute fact* but, rather, depends on implementation by a metaphysical link between the facts and the reason. It would be natural to conceive of this metaphysical link as a causal nexus. Without the metaphysical link, the modal tie between reasons and facts, which is necessary for reliability, cannot obtain. The demand for an explanation of reliability, in other words, is not a demand for a *causal* explanation of reliability but, rather, for a *reductive* explanation, which would describe the mechanism that realizes this reliability. Such a mechanism normally consists in a causal link between reasons and facts. In his explication of the explanationist objection, BonJour uses a striking example: "suppose that there is a person who holds a belief that is at least putatively about some specifiable element or region of reality, for reasons or evidence that seem initially substantial and compelling, but where neither the specific content of the belief nor the person's reasons for holding it are in fact causally shaped or otherwise influenced, directly or indirectly, by the element or region of reality in question. In such a situation, though the belief might still be true, it seems clear that its truth could only be accidental, a cognitive coincidence" (BonJour 1998: 157). But why does the reliability require implementation by a causal link? BonJour is perfectly clear on this point as well: "in the absence of such (causal) influence, the character of the reality in question could just as well have been different in such a way as to make the belief false without either the belief or its supporting reasons being affected in any way" (ibid.). So, as long as the normal causes of our reasons and beliefs lie elsewhere than in the truth-makers, the counterfactual co-variation of facts and reasons that would be required for reliability is missing.[8] If our reasons were not caused by the facts, the facts could have been different without our reasons registering that difference. And that undermines their reliability. This is exactly the situation in the case of clairvoyance. Which is why we do not think that clairvoyance leads to justified beliefs, regardless of how often any given clairvoyant may de facto be right in her predictions.

This line of thought makes the assumption plausible that there can be no reliable reasons that are not causally dependent upon the corresponding truth-makers. Thus, if we should discover that there is no causal link between truth-makers and reasons—or, indeed, as in the case of modal intuitions, that there cannot be any such link—that would seem to impugn the reliability of our reasons. We would have an undercutting defeater that would undermine the justificatory force of our reasons.

[8] See for a similar argument also Goldman / Pust 1998: 181 „if it is known or suspected that there is no relevant causal route or counterfactual dependence, there are grounds for doubting the existence of a reliable indicatorship relation".

So the slogan “There is no reliability unless the truth-makers have a causal influence on our reasons” may well be right in many cases. In my view, though, it cannot be generalized as it is in the argument for (P1). To see why, let’s take a look at our reasons for contingent truths. In this domain, reliability requires a metaphysical link between reasons and truth-makers. This link, however, can be realized in a number of different ways. First of all, the truth-makers could cause our reasons on a regular basis (as in the case of perception). Or the modal co-variance of reasons and truth-makers that is necessary for reliability could be explained by saying that the truth-makers supervene on facts, which, in turn, cause our reasons. This may be true for our moral intuitions, given that moral facts supervene on physical facts and are themselves not causally efficacious. Or the required co-variance could be explained by saying that our reasons cause the truth-making facts. Or, to put a provisional end to this list, the requisite co-variance could be explained by postulating common causes of our reasons and the truth-making facts. Even if consciousness is an epiphenomenon, as some suspect, there could still be a reliable link between our phenomenal judgments and the facts of consciousness, as long as they have a common causal pre-history. In short: the reductionist explanation of reliability does not have to imply that the truth-makers are the causes, and our reasons and beliefs the effects.

But it is a completely different matter if—as with our philosophical intuitions—our reasons support beliefs about modal reality. In that case, the modal tie between our reasons and the truth-makers, which is essential for reliability, can obtain even if it cannot be explained by any influence, interaction or relation of dependence between them. What BonJour prophesies for the event that such an influence should be absent—namely, “that in the absence of such influence, the character of the reality in question could just as well have been different in such a way as to make the belief false without either the belief or its supporting reasons being affected in any way”—cannot happen here, since modal reality, viewed counterfactually, cannot vary anyway. It is necessarily as it is. The stability of modal facts alone attests that the facts could not change without our reasons changing accordingly; indeed, the modal facts cannot change. Hence, it would be superfluous for the facts to influence our reasons. The reliability of our modal intuitions is simply a byproduct of the cognitive mechanism of our intuitions and the modal facts. There neither is nor needs to be a further explanation that describes metaphysical relations between the modal facts and our psychological mechanism. (See also Pust 2004) From the isolated standpoint of the laws of nature and from the isolated standpoint of the nature of modal reality, the veridicality of our modal intuitions seems to be purely accidental, because the two regions are metaphysically completely independent of each other. But if one takes account of both regions together, the modal link is every bit as tight as the causal nexus that connects perceptions and their distal causes. That can be seen clearly by considering the following point: if modal reality is not itself responsible for the occurrence of our intuitions about it, then, of course, those intuitions that are in fact reliable could lead us systematically astray if, for example, the initial conditions or the laws of the natural world were different leading to completely different outputs of the faculty of intuitions. But, if that were the case, the causal link between our perceptual experiences and the corresponding distal causes would also disappear. The impression that the relationship is accidental in the one case but not in the other rests upon a perspectival deception. The explanatory asymmetry that naturalists have observed between our natural cognitive abilities and philosophical intuitions does not point to a shortcoming of intuition but, rather, to a difference in kind between the two domains. The absence of an explanation is therefore no deficit, but a reflection of the fact that the reliability of our modal intuitions does not depend upon a metaphysical link between reasons and truth-makers. To construct from this an objection against the justificatory force of our intuitions would be to overlook the categorical difference between the two domains. If I am right about this, then (P1) is the result of a hasty over-generalization.

A proponent of the explanationist objection could perhaps be tempted to insist that there is at least a fundamental difference between, let's say, perception on the one hand, and modal intuition on the other, and that it is this difference that undermines the justificatory force of intuition in contrast to perception. For we can explain causally why organisms have reliable perceptual abilities with respect to their environments. Empirical evolutionary biology does so by noting that organisms without reliable perception cannot successfully negotiate their environments and thus fall victim to natural selection. There is no analogous causal explanation of why we should have *reliable* modal intuitions. That we can be prepared for various alternative situations is surely useful in planning how to act with limited knowledge of the actual future course of the world. But this usefulness depends only on our being able to pick out the actual future course of the world from all the possibilities under consideration, not on our being able to grasp non-actual possibilities correctly. There is no naturalistic causal explanation of why we have acquired reliable methods of modal knowledge. And, for reasons I have already mentioned, there cannot be a non-naturalistic causal explanation of this fact either. So the naturalist's observation is absolutely right: there is no causal explanation of why we have come to possess a reliable source of modal knowledge. But I do not see why any negative epistemological consequences should result from this really existing explanatory gap. As long as proponents of the explanationist objection leave this question unanswered, I regard this version of their objection as harmless.

I shall recapitulate briefly. I have tried to show two things: first, that philosophical intuitions can be understood as a priori evidence for modal judgments; and secondly, that the explanationist objection against the epistemic value of intuitions fails since it relies on at least one premise without good reason.[9]

REFERENCES

Alston, W. (1993). *The Reliability of Sense Perception*, Ithaca/London.

Bealer, G. (1998). 'Intuition and the Autonomy of Philosophy', *Rethinking Intuition* (eds.) DePaul & Ramsey, Lanham, 201–239.

— (2002). 'Modal Epistemology and Rationalism', *Conceivability and Possibility*, (eds.) T. Gendler & J. Hawthorne, Oxford, 71–125.

Benacerraf, P. (1973). 'Mathematical Truth', *The Journal of Philosophy* 70, 661–679.

BonJour, L. (1998). *In Defense of Pure Reason*, Cambridge.

Casullo, A. (2003). *A Priori Justification*, Oxford.

Field, H. (1989). *Realism, Mathematics and Modality*, Oxford.

Goldman, A. (1979). 'What is Justified Belief?', *Justification and Knowledge*, (ed.) G. Pappas, Dordrecht.

— (1992). 'Cognition and Modal Metaphysics', *Liaisons*, MIT, 49–66.

Goldman, A. & Pust, J. (1998). 'Philosophical Theory and Intuitional Evidence', *Rethinking Intuition*, (eds.) M. DePaul & W. Ramsey, Lanham, 179–197.

Gopnik, A. & Schwitzgebel, E. (1998). 'Whose concepts are they, anyway? The Role of Philosophical Intuition in Empirical Psychology', *Rethinking Intuition*, (eds.) M. DePaul & W. Ramsey, Lanham, 75–91.

[9] I am very grateful to Alvin Goldman, Frank Hofmann, Joachim Horvath, Ernest Sosa and Woldai Wagner for helpful comments on earlier drafts of this paper.

Kornblith, H. (1998). 'The Role of Intuition in Philosophical Inquiry: An Account with no Unnatural Ingredients', DePaul & Ramsey, 129–141.

Lewis, D. (1986). *On the Plurality of Worlds*, Oxford.

Pust, J. (2000). *Intuitions as Evidence*, New York & London.

— (2004). 'On Explaining Knowledge of Necessity', *Dialectica* 58, 71–87.

Sosa, E. (1998). 'Minimal Intuition', DePaul & Ramsey, 257–269.

— (Ms.). 'Intuitions'.

— (this volume). 'Intuitions: Their Nature and epistemic efficacy', *Grazer Philosophische Studien* 74, 51–67.

Williamson, T. (2004). 'Philosophical "Intuitions" and Scepticism about Judgement', *Dialectica* 58, 109–153.

— (2005). 'Armchair Philosophy, Metaphysical Modality and Counterfactual Thinking', *Proceedings of the Aristotelian Society* 105, 1–23.

Grazer Philosophische Studien
74 (2007), 89–123.

PHILOSOPHICAL KNOWLEDGE AND KNOWLEDGE OF COUNTERFACTUALS

Timothy WILLIAMSON
University of Oxford

Summary

Metaphysical modalities are definable from counterfactual conditionals, and the epistemology of the former is a special case of the epistemology of the latter. In particular, the role of conceivability and inconceivability in assessing claims of possibility and impossibility can be explained as a special case of the pervasive role of the imagination in assessing counterfactual conditionals, an account of which is sketched. Thus scepticism about metaphysical modality entails a more far-reaching scepticism about counterfactuals. The account is used to question the significance of the distinction between *a priori* and *a posteriori* knowledge.

§ 0.

Philosophers characteristically ask not just whether things are some way but whether they could have been otherwise. What could have been otherwise is *metaphysically contingent*; what could not is *metaphysically necessary*. We have some knowledge of such matters. We know that Henry VIII could have had more than six wives, but that three plus three could not have been more than six. So there should be an epistemology of metaphysical modality.

The differences between metaphysical necessity, contingency and impossibility are not mind-dependent, in any useful sense of that tantalizing phrase. Thus they are not differences in actual or potential psychological, social, linguistic or even epistemic status (Kripke 1980 makes the crucial distinctions). One shortcut to this conclusion uses the plausible idea that mathematical truth is mind-independent. Since mathematics is not contingent, the difference between truth and falsity in mathematics is also the difference between necessity and impossibility; consequently, the difference between necessity and impossibility is mind-independent. The difference between contingency and non-contingency is equally mind-independent;

for if C is a mind-independently true or false mathematical conjecture, then one of C and its negation conjoined with the proposition that Henry VIII had six wives forms a contingently true conjunction while the other forms an impossible conjunction, but which is which is mind-independent. To emphasize the point, think of the mind-independently truth-valued conjecture as evidence-transcendent, absolutely undecidable, neither provable nor refutable by any means. Thus the epistemology of metaphysical modality is one of mind-independent truths.

Nevertheless, doubts begin to arise. Although philosophers attribute metaphysical necessity to mathematical theorems, what matters mathematically is just their truth, not their metaphysical necessity: mathematics does not need the concept of metaphysical necessity. Does metaphysical modality really matter outside philosophy? Even if physicists care about the physical necessity of the laws they conjecture, does it matter to physics whether physically necessary laws are also metaphysically necessary? In ordinary life, we care whether someone could have done otherwise, or whether disaster could have been averted, but the kind of possibility at issue there is far more narrowly circumscribed than metaphysical possibility, by not prescinding from metaphysically contingent initial conditions. He could not have done otherwise because he was in chains, even though it was metaphysically contingent that he was in chains. Does "could have been" ever express metaphysical possibility when used non-philosophically?

If thought about metaphysical modality is the exclusive preserve of philosophers, so is knowledge of metaphysical modality. The epistemology of metaphysical modality tends to be treated as an isolated case. For instance, much of the discussion concerns how far, if at all, conceivability is a guide to possibility, and inconceivability to impossibility (Gendler and Hawthorne 2002 has a sample of recent contributions to this debate). The impression is that, outside philosophy, the primary cognitive role of conceiving is propaedeutic. Conceiving a hypothesis is getting it onto the table, putting it up for serious consideration as a candidate for truth. The inconceivable never even gets that far. Conceivability is certainly no good evidence for the restricted kinds of possibility that we care about in natural science or ordinary life. We easily conceive particles violating what are in fact physical laws, or the man without his chains. On this view, conceiving, outside philosophy, is not a faculty for distinguishing between truth and falsity in some domain, but rather a preliminary to any such faculty. Although there are truths and falsehoods about conceivability and inconceivability, they concern our mental capacities, whereas metaphysical modalities are supposed to be mind-independent. They are not contingent on mental capacities, because not contingent on anything (at least if we accept the principles of the modal logic S5, that the necessary is necessarily necessary and the possible necessarily possible). When philosophers present conceiving as a faculty for distinguishing between truth and falsity in the domain of metaphysical modality, that looks suspiciously like some sort of illicit projection or unacknowledged fiction: at best, attributions of metaphysical modality would lack the cognitive status traditionally ascribed to them (compare Blackburn 1987; Craig 1985; Wright 1989). The apparent cognitive isolation of metaphysically modal thought makes such suspicions hard to allay. Presenting it as *sui generis* suggests that it can be surgically removed from our conceptual scheme without collateral damage. If it can, what good does it do us? In general, the postulation by philosophers of a special cognitive capacity exclusive to philosophical or quasi-philosophical thinking looks like a scam.

Humans evolved under no pressure to do philosophy. Presumably, survival and reproduction in the stone age depended little on philosophical prowess (dialectical skill was probably no more effective then as a seduction technique than it is now). Any cognitive capacity that we have for philosophy is a more or less accidental byproduct of other developments. Nor are psychological dispositions that are non-cognitive outside philosophy likely suddenly to become cognitive within it. We should expect the cognitive capacities used in philosophy to be cases of general cognitive capacities used in ordinary life, perhaps trained, developed and systematically applied in various special ways, just as the cognitive capacities that we use in mathematics and natural science are rooted in more primitive cognitive capacities to perceive, count, reason, discuss …. In particular, a plausible non-sceptical epistemology of metaphysical modality should subsume our capacity to discriminate metaphysical possibilities from metaphysical impossibilities under more general cognitive capacities used in ordinary life.

I will argue that the ordinary cognitive capacity to handle counterfactual conditionals carries with it the cognitive capacity to handle metaphysical modality. § 1 illustrates with examples our cognitive use of counterfactual conditionals. § 2 sketches the beginnings of an epistemology of such conditionals. § 3 explains how they subsume metaphysical modality. § 4 discusses some objections. § 5 briefly raises the relation between metaphysical possibility and the restricted kinds of possibility that seem more relevant to ordinary life. Philosophers' ascriptions of metaphysical modality are far more deeply rooted in our ordinary cognitive practices than most sceptics realize.

§ 1.

We start with a well-known example that proves the term "counterfactual conditional" misleading. As Alan Ross Anderson pointed out (1951: 37), a doctor might say:

(1) If Jones had taken arsenic, he would have shown just exactly those symptoms which he does in fact show.

Clearly, (1) can provide abductive evidence by inference to the best explanation for its antecedent (see Edgington 2003: 23–7 for more discussion):

(2) Jones took arsenic.

If further tests subsequently verify (2), they confirm the doctor's statement rather than in any way falsifying it or making it inappropriate. If we still call subjunctive conditionals like (1) "counterfactuals", the reason is not that they imply or presuppose the falsity of their antecedents.

Of course, what (2) explains is not the trivial necessary truth that Jones shows whatever symptoms he shows. What is contingent is that Jones shows exactly those symptoms which he does in fact show—he could have shown other symptoms, or none—and, given (1), (2) explains that contingent truth.

While (1) provides valuable empirical evidence, the corresponding indicative conditional does not (Stalnaker 1999: 71):

(1I) If Jones took arsenic, he shows just exactly those symptoms which he does in fact show.

We can safely assent to (1I) without knowing what symptoms Jones shows, since it holds whatever they are. Informally, (1) is non-trivial because it depends on a comparison between independently specified terms, the symptoms which Jones would have shown if he had taken arsenic and the symptoms which he does in fact show; by contrast, (1I) is trivial because it involves only a comparison of his symptoms with themselves. Thus the process of evaluating the “counterfactual” conditional requires something like two files, one for the actual situation, the other for the counterfactual situation, even if these situations turn out to coincide. No such cross-comparison of files is needed to evaluate the indicative conditional. Of course, when one evaluates an indicative conditional while disbelieving its antecedent, one must not confuse one’s file of beliefs with one’s file of judgments on the supposition of the antecedent, but that does not mean that cross-referencing from the latter file to the former can play the role that it did in the counterfactual case.

Since (1) constitutes empirical evidence, its truth was not guaranteed in advance. If Jones had looked suitably different, the doctor would have had to assert the opposite counterfactual conditional:

(3) If Jones had taken arsenic, he would not have shown just exactly those symptoms which he does in fact show.

From (3) we can deduce the falsity of its antecedent. For modus ponens is generally agreed to be valid for counterfactual conditionals. Thus (2) and (3) yield:

(4) Jones does not show just exactly those symptoms which he does in fact show.

Since (4) is obviously false, we can deny (2) given (3).

The indicative conditional corresponding to (3) is:

(3I) If Jones took arsenic, he does not show just exactly those symptoms which he does in fact show.

To assert (3I) would be like saying “If Jones took arsenic, pigs can fly”. Although a very confident doctor might assert (3I), on the grounds that Jones certainly did not take arsenic, that certainty may in turn be based on confidence in (3), and therefore on the comparison of actual and counterfactual situations.

Could a Bayesian account dispense with the counterfactual conditionals in favour of conditional probabilities? Consider the simple case in which we completely trust the doctor who asserts (1). Before the doctor speaks, we are certain what symptoms Jones shows but agnostic over the characteristic symptoms of arsenic poisoning. We want to update our probability for his having taken arsenic on evidence from the doctor, in Bayesian terms by conditionalizing on it. The doctor cannot simply tell us what probability to assign, because we may have further relevant evidence unavailable to the doctor, for example about Jones's character. We need the doctor to say something which we can use as evidence; (1) exactly fits the bill (of course, our evidence also includes the fact that the doctor asserted (1), but in the circumstances we can treat (1) itself as the relevant part of our evidence). It may even do better than a non-modal generalization such as "Jones showed exactly those symptoms which everyone who takes arsenic shows": for the symptoms may vary with bodily characteristics of the victim, and through long experience the doctor may be able to judge what symptoms Jones would have shown if he had taken arsenic without being able to articulate a suitable generalization. Any Bayesian account depends on an adequately varied stock of propositions to act as bearers of probability, as evidence or hypotheses. Sometimes that range has to include counterfactual conditionals.

We also use the notional distinction between actual and counterfactual situations to make evaluative comparisons:

(5) If Jones had not taken arsenic, he would have been in better shape than he now is.

Such counterfactual reflections facilitate learning from experience; one may decide never to take arsenic oneself. Formulating counterfactuals about past experience is empirically correlated with improved future performance in various tasks.[1]

Evidently, counterfactual conditionals give clues to causal connections. This point does not commit one to the ambitious programme of analysing causality in terms of counterfactual conditionals (Lewis 1973b, Collins, Hall and Paul 2004), or counterfactual conditionals in terms of causality (Jackson 1977). If the former programme succeeds, all causal thinking is counterfactual thinking; if the latter succeeds, all counterfactual thinking is causal thinking. Either way, the overlap is so large that we cannot have one without much of the other. It may well be over-optimistic to expect either necessary and sufficient conditions for causal statements in counterfactual terms or necessary and sufficient conditions for counterfactual

[1] The large empirical literature on the affective role of counterfactuals and its relation to learning from experience includes Kahneman and Tversky 1982, Roese and Olson 1993, 1995 and Byrne 2005.

statements in causal terms. Even so, counterfactuals surely play a crucial role in our causal thinking (see Harris 2000: 118–139 and Byrne 2005: 100–128 for some empirical discussion). Only extreme sceptics deny the cognitive value of causal thought.

At a more theoretical level, claims of nomic necessity support counterfactual conditionals. If it is a law that property P implies property Q, then typically if something were to have P, it would have Q. If we can falsify the counterfactual in a specific case, perhaps by using better-established laws, we thereby falsify that claim of lawhood. We sometimes have enough evidence to establish what the result of an experiment would be without actually doing the experiment: that matters in a world of limited resources.

Counterfactual thought is deeply integrated into our empirical thought in general. Although that consideration will not deter the most dogged sceptics about our knowledge of counterfactuals, it indicates the difficulty of preventing such scepticism from generalizing implausibly far, since our beliefs about counterfactuals are so well-integrated into our general knowledge of our environment. I proceed on the assumption that we have non-trivial knowledge of counterfactuals.

§ 2.

In discussing the epistemology of counterfactuals, I assume no particular theory of their compositional semantics, although I sometimes use the Stalnaker-Lewis approach for purposes of illustration and vividness. That evasion of semantic theory might seem dubious, since it is the semantics which determines what has to be known. However, we can go some way on the basis of our pretheoretical understanding of such conditionals in our native language. Moreover, the best developed formal semantic theories of counterfactuals use an apparatus of possible worlds or situations at best distantly related to our actual cognitive processing. While that does not refute such theories, which concern the truth-conditions of counterfactuals, not how subjects attempt to find out whether those truth-conditions obtain, it shows how indirect the relation between the semantics and the epistemology may be. When we come to fine-tune our epistemology of counterfactuals, we may need an articulated semantic theory, but at a first pass we can make do with some sketchy remarks about their epistemology while remaining neutral over their deep semantic analysis.

As for the psychological study of the processes underlying our assessment of counterfactual conditionals, it remains in a surprisingly undeveloped state, as recent authors have complained (Evans and Over 2004: 113–131).

Start with an example. You are in the mountains. As the sun melts the ice, rocks embedded in it are loosened and crash down the slope. You notice one rock slide into a bush. You wonder where it would have ended if the bush had not been there. A natural way to answer the question is by visualizing the rock sliding without the bush there, then bouncing down the slope. You thereby come to know this counterfactual:

(6) If the bush had not been there, the rock would have ended in the lake.

You could test that judgment by physically removing the bush and experimenting with similar rocks, but you know (6) even without performing such experiments. Semantically, the counterfactual about the past is independent of claims about future experiments (for a start, the slope is undergoing continual small changes).

Somehow, you came to know the counterfactual by using your imagination. That sounds puzzling if one conceives the imagination as unconstrained. You can imagine the rock rising vertically into the air, or looping the loop, or sticking like a limpet to the slope. What constrains imagining it one way rather than another?

You do not imagine it those other ways because your imaginative exercise is radically informed and disciplined by your perception of the rock and the slope and your sense of how nature works. The default for the imagination may be to proceed as "realistically" as it can, subject to whatever deviations the thinker imposes by brute force: here, the absence of the bush. Thus the imagination can in principle exploit all our background knowledge in evaluating counterfactuals. Of course, how to separate background knowledge from what must be imagined away in imagining the antecedent is Goodman's old, deep problem of cotenability (1955). For example, why don't we bring to bear our background knowledge that the rock did not go far, and imagine another obstacle to its fall? Difficult though the problem is, it should not make us lose sight of our considerable knowledge of counterfactuals: our procedures for evaluating them cannot be too wildly misleading.

Can the imaginative exercise be regimented as a piece of reasoning? We can undoubtedly assess some counterfactuals by straightforward reasoning. For instance:

(7) If twelve people had come to the party, more than eleven people would have come to the party.

We can deduce the consequent "More than eleven people came to the party" from the antecedent "Twelve people came to the party", and assert (7) on that basis. Similarly, it may be suggested, we can assert (6) on the basis of inferring its consequent "The rock ended in the lake" from the premise "The bush was not there", given auxiliary premises about the rock, the mountainside and the laws of nature.

At the level of formal logic, we have the corresponding plausible and widely accepted closure principle that, given a derivation of **C** from $\mathbf{B}_1$, …, $\mathbf{B}_n$, we can derive the counterfactual conditional **A** □→ **C** from the counterfactual conditionals **A** □→ $\mathbf{B}_1$, …., **A** □→ $\mathbf{B}_n$; in other words, the counterfactual consequences of a supposition **A** are closed under logical consequence (Lewis calls this "Deduction within Conditionals", 1986: 132). With the uncontroversial reflexivity principle **A** □→ **A**, it follows that, given a derivation of **C** from **A** alone, we can derive **A** □→ **C** from the null set of premises.

We cannot automatically extend the closure rule to the case of auxiliary premises, for since we can derive an arbitrary conclusion **C** from an arbitrary premise **A** with **C** as auxiliary premise, we could then derive **A** □→ **C** from the auxiliary premise **C** alone: but that is in effect the invalid principle that any truth is a counterfactual consequence of any supposition whatsoever. Auxiliary premises cannot always be copied into the scope of counterfactual suppositions (the problem of cotenability again).

Even with this caution, the treatment of the process by which we reach counterfactual judgments as inferential is problematic in several ways.

First, a technical problem: not every inference licenses us to assert the corresponding counterfactual, even when the inference is deductive and the auxiliary premises are selected appropriately. For the consequent of (1) is a logical truth (count it vacuously true if Jones shows no symptoms):

(8) Jones shows just exactly those symptoms which he does in fact show.

Thus (8) follows from any premises, including (2), the antecedent of (1);

but we cannot assert (1) on the basis of that trivial deduction alone, independently of *which* symptoms Jones does in fact show. This is related to Kaplan's point that the rule of necessitation fails in languages with terms such as "actually" (1989). The logical truth of (8) does not guarantee the logical truth, or even truth, of (9):

(9) It is necessary that Jones shows just exactly those symptoms which he does in fact show.

For it is contingent that Jones shows just exactly those symptoms which he does in fact show.[2] But let us assume that this technical problem can be solved by a restriction on the type of reasoning from antecedent to consequent that can license a counterfactual, and on the closure principle above, like the restriction on the type of reasoning that licenses the necessitation of its conclusion.

A more serious problem is that the putative reasoner may lack general-purpose cognitive access to the auxiliary premises of the putative reasoning. In particular, the required folk physics may be stored in the form of some analogue mechanism, perhaps embodied in a connectionist network, which the subject cannot articulate in propositional form. Normally, a subject who uses negation and derives a conclusion from some premises can at least entertain the negation of a given premise, whether or not they are willing to assert it, perhaps on the basis of the other premises and the negation of the conclusion. Our reliance on folk physics does not enable us to entertain its negation. This strains the analogy with explicit reasoning.

The third problem is epistemological. Normally, someone who believes a conclusion on the sole basis of deduction from some premises knows the conclusion only if they know the premises. As a universally generalized theory, folk physics is presumably strictly speaking false: its predictions are inaccurate in some circumstances. Consequently, it is not known. But the conclusion that no belief formed on the basis of folk physics constitutes knowledge is wildly sceptical. For folk physics is reliable enough in many circumstances to be used in the acquisition of knowledge, for example that the cricket ball will land in that field. Thus we should not conceive folk physics as a premise of that conclusion. Nor should we conceive some

[2] The phrase "does in fact show" is read throughout as inside the scope of the counterfactual conditional or modal operator, but as rigid, like "actually shows". See Williamson 2006 for relevant discussion.

local fragment of folk physics as the premise. For it would be quite unmotivated to take an inferential approach overall while refusing to treat this local fragment as itself derived from the general theory of folk physics. We should conceive folk physics as a locally but not globally reliable method of belief formation, not as a premise.

The preceding reasons motivate the attempt to understand the imaginative exercises by which we judge counterfactuals like (6) as not purely inferential. An attractive suggestion is that some kind of simulation is involved: the difficulty is to explain what that means. It is just a hint of an answer to say that in simulation cognitive faculties are run off-line. The cognitive faculties that would be run on-line to evaluate **A** and **B** as freestanding sentences are run off-line in the evaluation of the counterfactual conditional $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$.[3] This suggests that the cognition has a roughly compositional structure. Our capacity to handle $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$ embeds our capacities to handle **A** and **B**, and our capacity to handle the counterfactual conditional operator involves a general capacity to go from capacities to handle the antecedent and the consequent to a capacity to handle the whole conditional. Here the capacity to handle an expression generally comprises more than mere linguistic understanding of it, since it involves ways of assessing its application that are not built into its meaning. But it virtually never involves a decision procedure that enables us always to determine the truth-values of every sentence in which the expression principally occurs, since we lack such decision procedures. Of course, we can sometimes take shortcuts in evaluating counterfactual conditionals. For instance, we can know that $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$ is true even if we have no idea how to determine whether **A** is true. Nevertheless, the compositional structure just described seems more typical.

How do we advance from capacities to handle the antecedent and the consequent to a capacity to handle the whole conditional? "Off-line" suggests that the most direct links with perception have been cut, but that vague negative point does not take us far. Perceptual input is crucial to the evaluation of counterfactuals such as (1) and (6).

The best developed simulation theories concern our ability to simulate the mental processes of other agents (or ourselves in other circumstances), putting ourselves in their shoes, as if thinking and deciding on the basis of their beliefs and desires (see for example Davies and Stone 1995, Nichols

[3] Matters become more complicated if **A** or **B** itself contains a counterfactual condition, as in "If she had murdered the man who would have inherited her money if she had died, she would have been sentenced to life imprisonment if she had been convicted".

and Stich 2003). Such cognitive processes may well be relevant to the evaluation of counterfactuals about agents. Moreover, they would involve just the sort of constrained use of the imagination indicated above. How would Mary react if you asked to borrow her car? You could imagine her immediately shooting you, or making you her heir; you could even imagine reacting like that from her point of view, by imagining having sufficiently bizarre beliefs and desires. But you do not. Doing so would not help you determine how she really would react. Presumably, what you do is to hold fixed her actual beliefs and desires (as you take them to be just before the request); you can then imagine the request from her point of view, and think through the scenario from there. Just as with the falling rock, the imaginative exercise is richly informed and disciplined by your sense of what she is like.

How could mental simulation help us evaluate a counterfactual such as (6), which does not concern an agent? Even if you somehow put yourself in the rock's shoes, imagining first-personally being that shape, size and hardness and bouncing down that slope, you would not be simulating the rock's reasoning and decision-making. Thinking of the rock as an agent is no help in determining its counterfactual trajectory. A more natural way to answer the question is by imagining third-personally the rock falling as it would visually appear from your actual present spatial position; you thereby avoid the complex process of adjusting your current visual perspective to the viewpoint of the rock. Is that to simulate the mental states of an observer watching the rock fall from your present position?[4] By itself, that suggestion explains little. For how do we know what to simulate the observer seeing next? But that question is not unanswerable. For we have various propensities to form expectations about what happens next: for example, to project the trajectories of nearby moving bodies into the immediate future (otherwise we could not catch balls). Perhaps we simulate the initial movement of the rock in the absence of the bush, form an expectation as to where it goes next, feed the expected movement back into the simulation as seen by the observer, form a further expectation as to its subsequent movement, feed that back into the simulation, and so on. If our expectations in such matters are approximately correct in a range of ordinary cases, such a process is cognitively worthwhile. The very natural laws and causal tendencies which our expectations roughly track also help to determine which counterfactual conditionals really hold.

[4] See Goldman 1992: 24, discussed by Nichols, Stich, Leslie and Klein 1996: 53–59.

However, talk of simulating the mental states of an observer may suggest that the presence of the observer is part of the content of the simulation. That does not fit our evaluation of counterfactuals. Consider:

(10) If there had been a tree on this spot a million years ago, nobody would have known.

Even if we visually imagine a tree on this spot a million years ago, we do not automatically reject (10) because we envisage an observer of the tree. We may imagine the tree as having a certain visual appearance from a certain viewpoint, but that is not to say that we imagine it as appearing to someone at that viewpoint. For example, if we imagine the sun as shining from behind that viewpoint, by imagining the tree's shadow stretching back from the tree, we are not obliged to imagine either the observer's shadow stretching towards the tree or the observer as perfectly transparent.[5] Nor, when we consider (10), are we asking whether if we had believed that there was a tree on this spot a million years ago, we would have believed that nobody knew.[6] It may be better not to think of the simulation as specifically *mental* simulation at all.

Of course, for many counterfactuals the relevant expectations are not hardwired into us in the way that those concerning the trajectories of fast-moving objects around us may need to be. Our knowledge that if a British general election had been called in 1948 the Communists would not have won may depend on an off-line use of our capacity to predict political events. Still, where our more sophisticated capacities to predict the future are reliable, so should be corresponding counterfactual judgments. In these cases too, simulating the mental states of an imaginary observer seems unnecessary.

[5] The question is of course related to Berkeley's claim that we cannot imagine an unseen object. For discussion see Williams 1966, Peacocke 1985 and Currie 1995: 36–37.

[6] A similar problem arises for what is sometimes called the Ramsey Test for conditionals, on which one simulates belief in the antecedent and asks whether one then believes the consequent. Goldman writes "When considering the truth value of "If X were the case, then Y would obtain," a reasoner feigns a belief in X and reasons about Y under that pretense" (1992: 24). What Ramsey himself says is that when people "are fixing their degrees of belief in *q* given *p*" they "are adding *p* hypothetically to their stock of knowledge and arguing on that basis about *q*" (1978: 143), but he specifically warns that "the degree of belief in *q* given *p*" does not mean the degree of belief "which the subject would have in *q* if he knew *p*, or that which he ought to have" (1978: 82; variables interchanged). Of course, conditional probabilities bear more directly on indicative than on subjunctive conditionals.

The off-line use of expectation-forming capacities to judge counterfactuals corresponds to the widespread picture of the semantic evaluation of those conditionals as "rolling back" history to shortly before the time of the antecedent, modifying its course by stipulating the truth of the antecedent and then rolling history forward again according to patterns of development as close as possible to the normal ones to test the truth of the consequent (compare Lewis 1979). Not all counterfactual conditionals can be so evaluated, since the antecedent need not concern a limited time: in evaluating the claim that space-time has ten dimensions, a scientist can sensibly ask whether if it were true the actually observed phenomena would have occurred. Explicit reasoning may play a much larger role in the evaluation of such conditionals.

Reasoning and prediction do not exhaust our capacity to evaluate counterfactuals. If twelve people had come to the party, would it have been a large party? To answer, one does not imagine a party of twelve people and then predict what would happen next. The question is whether twelve people would have constituted a large party, not whether they would have caused one. Nor is the process of answering best conceived as purely inferential, if one has no special antecedent beliefs as to how many people constitute a large party, any more than the judgment whether the party is large is purely inferential when made at the party. Rather, in both cases one must make a new judgment, even though it is informed by what one already believes or imagines about the party. To call the new judgment "inferential" simply because it is not made independently of all the thinker's prior beliefs or suppositions is to stretch the term "inferential" beyond its useful span. At any rate, the judgment cannot be derived from the prior beliefs or suppositions purely by the application of general rules of inference. For example, even if you have the prior belief that a party is large if and only if it is larger than the average size of a party, in order to apply it to the case at hand you also need to have a belief as to what the average size of a party is; if you have no prior belief as to that, and must form one by inference, an implausible regress threatens, for you do not have the statistics of parties in your head. Similarly, if you try to judge whether this party is large by projecting inductively from previous judgments as to whether parties were large, that only pushes the question back to how those previous judgments were made.

In general, our capacity to evaluate counterfactuals recruits *all* our cognitive capacities to evaluate sentences. A quick proof of this uses the assumption that a counterfactual with a true antecedent has the same truth-value as its consequent, for then any sentence $\mathbf{A}$ is logically equivalent to $\mathbf{T} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$, where $\mathbf{T}$ is a trivial tautology; so any serious cognitive work needed to evaluate $\mathbf{A}$ is also needed to evaluate $\mathbf{T} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$.[7]

We can schematize the process of evaluating a counterfactual conditional thus: the thinker imaginatively supposes the antecedent and counterfactually develops the supposition, adding further judgments within the supposition by reasoning, off-line predictive mechanisms and other off-line judgments. To a first approximation: if the development eventually leads us to add the consequent, we assent to the conditional; if not, we dissent from it. Of course, this initial sketch is much too crude, in several ways. We may not be confident enough about the background conditions to decide for or against the conditional. Even if we are confident enough in that respect, if the consequent has not emerged after a given period of development the question remains whether it will emerge in the course of further development, for lines of reasoning can be continued indefinitely from any given premise. To reach a negative conclusion, we must in effect judge that if the consequent were ever going to emerge it would have done so by now (for example, we may have been smoothly fleshing out a scenario incompatible with the consequent with no hint of difficulty). A further over-simplification was that we develop the initial supposition only once: if we find various different ways of imagining the antecedent holding equally good, we may try developing several of them, to see whether they all yield the consequent. For example, if in considering (10) you initially imagine a palm tree, you do not immediately judge that if there had been a tree on this spot a million years ago it would have been a palm tree, because you know that you can equally easily imagine a fir tree. Although far more needs to be said, these remarks may at least start us in the right direction.

Despite its discipline, our imaginative evaluation of counterfactual conditionals is manifestly fallible. We can easily misjudge their truth-values, through background ignorance or error, and distortions of judgment. But such fallibility is the common lot of human cognition. Our use of the imagination in evaluating counterfactuals is practically indispensable. Rather than cave in to scepticism, we should admit that our methods sometimes yield knowledge of counterfactuals.

[7] Lewis defends the assumption (1986: 26–31); Nozick rejects it to make the fourth condition in his analysis of knowledge non-trivial (1981: 176). Bennett also rejects it (2003: 239–40).

§3.

How does the epistemology of counterfactual conditionals bear on the epistemology of metaphysical modality? We can approach this question by formulating two plausible constraints on the relation between counterfactual conditionals and metaphysical modalities. Henceforth, "necessary" and "possible" will be used for the metaphysical modalities unless otherwise stated.

First, the strict conditional implies the counterfactual conditional:

NECESSITY $\Box(\mathbf{A} \supset \mathbf{B}) \supset (\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B})$

Suppose that $\mathbf{A}$ could not have held without $\mathbf{B}$ holding too; then if $\mathbf{A}$ had held, $\mathbf{B}$ would also have held. In terms of possible worlds semantics for these operators along the lines of Lewis (1973) or Stalnaker (1968): if all $\mathbf{A}$ worlds are $\mathbf{B}$ worlds, then any closest $\mathbf{A}$ worlds are $\mathbf{B}$ worlds. More precisely, if all $\mathbf{A}$ worlds are $\mathbf{B}$ worlds, then either there are no $\mathbf{A}$ worlds or there is an $\mathbf{A}$ world such that any $\mathbf{A}$ world at least as close as it is to the actual world is a $\mathbf{B}$ world.

Second, the counterfactual conditional transmits possibility:

POSSIBILITY $(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}) \supset (\Diamond\mathbf{A} \supset \Diamond\mathbf{B})$

Suppose that if $\mathbf{A}$ had held, $\mathbf{B}$ would also have held; then if it is possible for $\mathbf{A}$ to hold, it is also possible for $\mathbf{B}$ to hold. In terms of worlds: if any closest $\mathbf{A}$ worlds are $\mathbf{B}$ worlds, and there are $\mathbf{A}$ worlds, then there are also $\mathbf{B}$ worlds. More precisely, if either there are no $\mathbf{A}$ worlds or there is an $\mathbf{A}$ world such that any $\mathbf{A}$ world at least as close as it is to the actual world is a $\mathbf{B}$ world, then if there is an $\mathbf{A}$ world there is also a $\mathbf{B}$ world.

Together, NECESSITY and POSSIBILITY sandwich the counterfactual conditional between two modal conditions. But they do not squeeze it very tight, for $\Diamond\mathbf{A} \supset \Diamond\mathbf{B}$ is much weaker than $\Box(\mathbf{A} \supset \mathbf{B})$: although the latter entails the former in any normal modal logic, the former is true and the latter false whenever $\mathbf{B}$ is possible without being a necessary consequence of $\mathbf{A}$, for example when $\mathbf{A}$ and $\mathbf{B}$ are modally independent.

Although NECESSITY and POSSIBILITY determine no necessary and sufficient condition for the counterfactual conditional in terms of necessity and possibility, they yield necessary and sufficient conditions for necessity and possibility in terms of the counterfactual conditional.

We argue thus. Let ⊥ be a contradiction. As a special case of NECESSITY:

(11) $\Box(\neg \mathbf{A} \supset \bot) \supset (\neg \mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot)$

By elementary (normal) modal logic, since a truth-functional consequence of something necessary is itself necessary:

(12) $\Box \mathbf{A} \supset \Box(\neg \mathbf{A} \supset \bot)$

From (11) and (12) by transitivity of the material conditional:

(13) $\Box \mathbf{A} \supset (\neg \mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot)$

Similarly, as a special case of POSSIBILITY:

(14) $(\neg \mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot) \supset (\Diamond\neg \mathbf{A} \supset \Diamond \bot)$

By elementary (normal) modal logic, since the possibility of a contradiction is itself inconsistent, and necessity is the dual of possibility (being necessary is equivalent to having an impossible negation):

(15) $(\Diamond\neg \mathbf{A} \supset \Diamond \bot) \supset \Box \mathbf{A}$

From (14) and (15) by transitivity:

(16) $(\neg \mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot) \supset \Box \mathbf{A}$

Putting (13) and (16) together:

(17) $\Box \mathbf{A} \equiv (\neg \mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot)$

The necessary is that whose negation counterfactually implies a contradiction. Since possibility is the dual of necessity (being possible is equivalent to having an unnecessary negation), (17) yields a corresponding necessary and sufficient condition for possibility, once a double negation in the antecedent of the counterfactual has been eliminated.

(18) $\Diamond \mathbf{A} \equiv \neg(\mathbf{A} \mathrel{\Box\!\!\rightarrow} \bot)$

The impossible is that which counterfactually implies a contradiction; the possible is that which does not. In (17) and (18), the difference between necessity and possibility lies simply in the scope of negation.

Without assuming a specific framework for the semantics of counterfactuals (in particular, that of possible worlds), we can give a simple semantic rationale for (17) and (18), based on the idea of vacuous truth. That some true counterfactuals have impossible antecedents is clear, for otherwise $\mathbf{A} \Box\!\!\rightarrow \mathbf{A}$ would fail when $\mathbf{A}$ was impossible. Make two generally accepted assumptions about the distinction between vacuous and non-vacuous truth: (a) $\mathbf{B} \Box\!\!\rightarrow \mathbf{C}$ is vacuously true if and only if $\mathbf{B}$ is impossible (this could be regarded as a definition of "vacuously" for counterfactuals); (b) $\mathbf{B} \Box\!\!\rightarrow \mathbf{C}$ is non-vacuously true only if $\mathbf{C}$ is possible. The truth of (17) and (18) follows, given normal modal reasoning. If $\Box\mathbf{A}$ is true, then $\neg\mathbf{A}$ is impossible, so by (a) $\neg\mathbf{A} \Box\!\!\rightarrow \bot$ is vacuously true; conversely, if $\neg\mathbf{A} \Box\!\!\rightarrow \bot$ is true, then by (b) it is vacuously true, so by (a) $\neg\mathbf{A}$ is impossible, so $\Box\mathbf{A}$ is true. Similarly, if $\Diamond\mathbf{A}$ is true, then $\mathbf{A}$ is not impossible, so by (a) $\mathbf{A} \Box\!\!\rightarrow \bot$ is not vacuously true, and by (b) not non-vacuously true, so $\neg(\mathbf{A} \Box\!\!\rightarrow \bot)$ is true; if $\Diamond\mathbf{A}$ is not true, then $\mathbf{A}$ is impossible, so by (a) $\mathbf{A} \Box\!\!\rightarrow \bot$ is vacuously true, so $\neg(\mathbf{A} \Box\!\!\rightarrow \bot)$ is not true.

Given that the equivalences (17) and (18) are logically true, metaphysically modal thinking is logically equivalent to a special case of counterfactual thinking, and the epistemology of the former is tantamount to a special case of the epistemology of the latter. Whoever has what it takes to understand the counterfactual conditional and the elementary logical auxiliaries $\neg$ and $\bot$ has what it takes to understand possibility and necessity operators.

The definability of necessity and possibility in terms of counterfactual conditionals was recognized long ago. It is easy to show from the closure and reflexivity principles for $\Box\!\!\rightarrow$ in § 2 that $\mathbf{A} \Box\!\!\rightarrow \bot$ is logically equivalent to $\mathbf{A} \Box\!\!\rightarrow \neg\mathbf{A}$. Thus (17) and (18) generate two new equivalences:

(19) $\Box\mathbf{A} \equiv (\neg\mathbf{A} \Box\!\!\rightarrow \mathbf{A})$

(20) $\Diamond\mathbf{A} \equiv \neg(\mathbf{A} \Box\!\!\rightarrow \neg\mathbf{A})$

The necessary is that which is counterfactually implied by its own negation; the possible is that which does not counterfactually imply its own negation. Stalnaker (1968) used (19) and (20) to define necessity and possibility, although his reading of the conditional (with a different notation) was not exclusively counterfactual. Lewis (1973a: 25) used (17) and (18) themselves to define necessity and possibility in terms of the counterfactual conditional. However, such definitions seem to have been treated as convenient notational economies, their potential philosophical significance unnoticed (Hill 2006 is a recent exception).

If we permit ourselves to quantify into sentence position ("propositional quantification"), we can formulate another pair of variants on (17) and (18) that may improve our feel for what is going on.[8] On elementary assumptions about the logic of such quantifiers and of the counterfactual conditional, $\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$ is provably equivalent to $\forall \boldsymbol{p}\ (\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \mathbf{A})$: something is counterfactually implied by its negation if and only if it is counterfactually implied by everything. Thus (19) and (20) generate these equivalences too:

(21) $\Box\mathbf{A} \equiv \forall \boldsymbol{p}\ (\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \mathbf{A})$

(22) $\Diamond\mathbf{A} \equiv \exists \boldsymbol{p}\ \neg(\boldsymbol{p} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{A})$

According to (21), something is necessary if and only if whatever were the case, it would still be the case (see also Lewis 1986: 23). That is a natural way of explaining informally what metaphysically necessity is. According to (22), something is possible if and only if it is not such that it would fail in every eventuality.

Since the right-hand sides of (17), (19) and (21) are not strictly synonymous with each other, given the differences in their semantic structure, they are not all strictly synonymous with $\Box\mathbf{A}$. Similarly, since the right-hand sides of (18), (20) and (22) are not strictly synonymous with each other, they are not all strictly synonymous with $\Diamond\mathbf{A}$. Indeed, we have no sufficient reason to regard any of the equivalences as strict synonymies. That detracts little from their philosophical significance, for failure of strict synonymy does not imply failure of logical equivalence. The main philosophical concerns about possibility and necessity apply equally to anything logically equivalent to possibility or necessity. A non-modal analogy:

[8] This quantification into sentence position need not be understood substitutionally. In purely modal contexts it can be modeled as quantification over all sets of possible worlds, even if not all of them are intensions of sentences that form the supposed substitution class, although this modeling presumably fails for hyperintensional contexts such as epistemic ones. A more faithful semantics for it might use non-substitutional quantification into sentence position in the meta-language. Such subtleties are inessential for present purposes.

$\neg$**A** is logically equivalent to **A** $\supset \bot$, but presumably they are not strictly synonymous; nevertheless, once we have established that a creature can handle $\supset$ and $\bot$, we have established that it can handle something logically equivalent to negation, which answers the most interesting questions about its ability to handle negation. We should find the mutual equivalence of (17), (19) and (21), and of (18), (20) and (22) reassuring, for it shows the robustness of the modal notions definable from the counterfactual conditional, somewhat as the equivalence of the various proposed definitions of "computable function" showed the robustness of that notion.

If we treat (17) and (18) like definitions of $\Box$ and $\Diamond$ for logical purposes, and assume some elementary principles of the logic of counterfactuals, then we can establish the main principles of elementary modal logic for $\Box$ and $\Diamond$. For example, we can show that what follows from necessary premises is itself necessary. Given that counterfactual conditionals obey modus ponens (or even weaker assumptions), we can show that what is necessary is the case. We can also check that the principles NECESSITY and POSSIBILITY, which we used to establish (17) and (18), do indeed hold under the latter characterizations of necessity and possibility. Under much stronger assumptions about the logic of the counterfactual conditional, we can also establish much stronger principles of modal logic, such as the S5 principle that what is possible is necessarily possible. Such connections extend to quantified modal logic. The logic of counterfactual conditionals smoothly generates the logic of the modal operators. Technical details are omitted here.

In particular, the proposed conception of modality makes quantification into the scope of modal operators tantamount to a special case of quantification into counterfactual contexts, as in (23) and (24):

(23) Everyone who would have benefited if the measure had passed voted for it.

(24) Where would the rock have landed if the bush had not been there?

Thus challenges to the intelligibility of claims of *de re* necessity are tantamount to challenges to the intelligibility of counterfactuals such as (23) and (24). But (23) and (24) are evidently intelligible.

Given (17) and (18), we should expect the epistemology of metaphysical modality to be a special case of the epistemology of counterfactuals. Far from being *sui generis*, the capacity to handle metaphysical modality will be an "accidental" byproduct of the cognitive mechanisms which provide our capacity to handle counterfactual conditionals. Since our capacity for modal thinking cannot be isolated from our capacity for ordinary thinking about the natural world, which involves counterfactual thinking, sceptics cannot excise metaphysical modality from our conceptual scheme without loss to ordinary thought about the natural world, for the former is implicit in the latter.

A useful comparison is with the relation between logical consequence and logical truth. Consider some agents who reason in simple ways about themselves and their environment, perhaps using rules of inference formalizable in a Gentzen-style natural deduction calculus, perhaps in some less sophisticated way. The practical value of their reasoning skill is that they can move from ordinary empirical premises to ordinary empirical conclusions in ways that always preserve truth, thereby extending their knowledge of mundane matters (see Schechter 2006 for relevant discussion). In doing so, they need never use logically true sentences. Nevertheless, the cognitive capacity that enables them to make these transitions between empirical sentences also enables them, as a special case, an "accidental" byproduct, to deduce logical truths from the null set of premises. Highly artificial moves would be needed to block these bonus deductions; such *ad hoc* restrictions would come at the price of extra computational complexity for no practical gain. Likewise at the semantic level: The simplest compositional semantics that enables us to negate and conjoin empirical sentences also enables us to formulate logical truths and falsehoods, even if we have hitherto lacked any interest in doing so. By good fortune, everything is already in place for the logician to evaluate logical truths and falsehoods (at least in first-order logic, since it is complete). The philosopher's position with respect to metaphysical modality is not very different.

Discussions of the epistemology of modality often focus on imaginability or conceivability as a test of possibility while ignoring the role of the imagination in the assessment of mundane counterfactuals. In doing so, they omit the appropriate context for understanding the relation between modality and the imagination. For instance, scorn is easily poured on imagination as a test of possibility: it is imaginable but not possible that water does not contain oxygen, except in artificial senses of "imaginable" that come apart from possibility in other ways, and so on. Imagination can be made to look cognitively worthless. Once we recall its fallible but vital role in evaluating counterfactual conditionals, we should be more open to the idea that it plays such a role in evaluating claims of possibility and necessity. At the very least, we cannot expect an adequate account of the role of imagination in the epistemology of modality if we lack an adequate account of its role in the epistemology of counterfactuals.

On the simplest version of the account in § 2, we accept $\mathbf{A} \Box\!\!\rightarrow \mathbf{B}$ when our counterfactual development of the supposition **A** generates **B**; we reject $\mathbf{A}\Box\!\!\rightarrow\mathbf{B}$ when our counterfactual development of **A** fails to generate **B** (in a reasonable time). Thus, by (17), we accept $\Box\mathbf{A}$ when our counterfactual development of the supposition $\neg\mathbf{A}$ generates a contradiction; we reject $\Box\mathbf{A}$ when our counterfactual development of $\neg\mathbf{A}$ fails to generate a contradiction (in a reasonable time). Similarly, by (18), we accept $\Diamond\mathbf{A}$ when our counterfactual development of the supposition **A** fails to generate a contradiction (in a reasonable time); we reject $\Diamond\mathbf{A}$ when our counterfactual development of **A** generates a contradiction. Thus our fallible imaginative evaluation of counterfactuals has a conceivability test for possibility and an inconceivability test for impossibility as fallible special cases. Such conceivability and inconceivability will be subject to the same constraints, whatever they are, as counterfactual conditionals in general, concerning which parts of our background information are held fixed. If we know enough chemistry, our counterfactual development of the supposition that gold is the element with atomic number 79 will generate a contradiction. The reason is not simply that we know that gold is the element with atomic number 79, for we can and must vary some items of our knowledge under counterfactual suppositions. Rather, general constraints on the development of counterfactual suppositions require us to hold such constitutive facts fixed.

A nuanced account of our handling of counterfactuals is likely to predict that we are more reliable in evaluating some kinds than others. For example, we may well be more reliable in evaluating counterfactuals whose antecedents involve small departures from the actual world than in evaluating those whose antecedents involve much larger departures. We may be correspondingly more reliable in evaluating the possibility of everyday scenarios than of "far-out" ones, and extra caution may be called for in the latter case. At the limit, actuality is often the best argument for possibility. But current philosophical practice already shows some sensitivity to such considerations. We may be more confident of the possibility of more or less realistic thought experiments in epistemology and moral philosophy than of more radically strange ones in metaphysics. More explicit consideration of the link between modal thought and counterfactual thought may lead to further refinements of our practice. But the use of imagination to evaluate philosophical claims of possibility and necessity is not illegitimate in principle, any more than is its use to evaluate mundane counterfactuals.

What does the envisaged assimilation of modality to counterfactual conditionals imply for the status of modal judgments as knowable *a priori* or only *a posteriori*? Some counterfactual conditions look like paradigms of *a priori* knowability: for example (7), whose consequent is a straightforward deductive consequence of its antecedent. Others look like paradigms of what can be known only *a posteriori*: for example, that if I had searched in my pocket five minutes ago I would have found a coin. But those are easy cases.

Standard discussions of the *a priori* distinguish between two roles that experience plays in cognition, one *evidential*, one *enabling*. Experience is held to play an evidential role in my visual knowledge that this shirt is green, but a merely enabling role in my knowledge that all green things are coloured: I needed it only to acquire the concepts *green* and *coloured*, without which I could not even raise the question whether all green things are coloured. Knowing *a priori* is supposed to be incompatible with an evidential role for experience, so my knowledge that this shirt is green is not *a priori*; but compatible with an enabling role for experience, so my knowledge that all green things are coloured can still be *a priori*. However, in our imagination-based knowledge of counterfactuals, experience can play a role that is neither strictly evidential nor purely enabling. For it can mould the ways in which we later imagine and judge, beyond what is needed to grasp the relevant concepts, without surviving as part of our total evidence.

Here is an example. I acquire the words “inch” and “centimetre” independently of each other. Through experience, I learn to make naked eye judgments of distances in inches or centimetres with moderate reliability. When things go well, such judgments amount to knowledge: *a posteriori* knowledge, of course. For example, I know *a posteriori* that two marks in front of me are at most two inches apart. Now I deploy the same faculty off-line to make a counterfactual judgment:

(25) If these marks had been at least nine inches apart, they would have been at least nineteen centimetres apart.

In judging (25), I do not use a conversion ratio between inches and centimetres to make a calculation. In the example I know no such ratio.

Rather, I visually imagine the two marks nine inches apart, and use my ability to judge distances in centimetres visually off-line to judge under the counterfactual supposition that the marks are at least nineteen centimetres apart. With this large margin for error, my judgment is reliable. Thus I know (25). Do I know it *a priori* or *a posteriori*? Experience plays no direct evidential role in my judgment. I do not consciously or unconsciously recall memories of distances encountered in perception, nor do I deduce (25) from general principles that I have inductively or abductively gathered from experience: § 2 noted obstacles to assimilating counterfactual thinking to reasoning. Nevertheless, the causal role of past experience in my judgment of (25) far exceeds enabling me to grasp the concepts in (25). Someone could easily have enough experience to understand (25) without being reliable enough in their judgments of distance to know (25).

If we classify my knowledge of (25) in the envisaged circumstances as *a priori*, because experience plays no strictly evidential role, the danger is that far too much will count as *a priori*. Experience can mould my judgment in many ways without playing a direct evidential role. But if we classify my knowledge of (25) as *a posteriori*, because experience plays more than a purely enabling role, that may apply to many philosophically significant modal judgments too. Of course, Kripke has argued strongly for a category of necessary truths knowable only *a posteriori*, such as "Gold is the element with atomic number 79"; "It is necessary that gold is the element with atomic number 79" would then be knowable only *a posteriori* too. The present suggestion is intended far more widely than that. For example:

(26) It is necessary that whoever knows something believes it.

(27) If Mary knew that it was raining, she would believe that it was raining.

Knowledge of truths such as (26) and (27) is usually regarded as *a priori*, even by those who accept the category of the necessary *a posteriori*. The experiences through which we learned to distinguish in practice between belief and non-belief and between knowledge and ignorance play no strictly evidential role in our knowledge of (26) and (27). Nevertheless, their role may be more than purely enabling. Many philosophers, native speakers of English, have denied (26) (Shope 1983: 171–192 has a critical survey). They are not usually or plausibly accused of failing to understand the words "know" and "believe". Why should not subtle differences between two courses of experience, each of which sufficed for coming to understand "know" and "believe", make for differences in how test cases are imagined, just large enough to tip honest judgments in opposite directions? Whether knowledge of (26) and (27) is available to one may thus be highly sensitive to personal circumstances.

If that picture is on the right lines, should we conclude that modal knowledge is *a posteriori*? Not if that suggests that (26) and (27) are inductive or abductive conclusions from perceptual data. In such cases, the question "*A priori* or *a posteriori*?" is too crude to be of much epistemological use. The point is not that we cannot draw a line somewhere with traditional paradigms of the *a priori* on one side and traditional paradigms of the *a posteriori* on the other. Surely we can; the point is that doing so yields little insight. The distinction is handy enough for a rough initial description of epistemic phenomena; it is out of place in a deeper theoretical analysis, because it obscures more significant epistemic patterns.[9]

§ 4.

It is time to consider objections to the preceding account.

Objection: Knowledge of counterfactuals cannot explain modal knowledge, because the former depends on the latter. More specifically, in developing a counterfactual supposition, we make free use of what we take to be necessary truths, but not of what we take to be contingent truths. Thus we rely on a prior stock of modal knowledge or belief. The principle NECESSITY above illustrates how we do this.

Reply: Once we take something to be a necessary truth, of course we can use it in developing further counterfactual suppositions. But that does nothing to show that we have any special cognitive capacity to handle modality independent of our general cognitive capacity to handle counterfactual conditionals. If we start only with the latter, just as envisaged above, it will generate knowledge of various modal truths, which can in turn be used to develop further counterfactual suppositions, in a recursive process. For example, we need not judge that it is metaphysically necessary that gold is the element with atomic number 79 *before* invoking the proposi-

[9] This problem for the *a priori*/*a posteriori* distinction undermines arguments for the incompatibility of semantic externalism with our privileged access to our own mental states that appeal to the supposed absurdity of *a priori* knowledge of contingent features of the external environment (McKinsey 1991).

tion that gold is the element with atomic number 79 in the development of a counterfactual supposition. Rather, projecting constitutive matters such as atomic numbers into counterfactual suppositions is part of our general way of assessing counterfactuals. The judgment of metaphysical necessity originates as the output of a procedure of that kind; it is not an independently generated input.

Objection: The account associates metaphysical modality with counterfactual conditionals of a very peculiar kind: in the case of (17) and (18), those with an explicit contradiction as their consequent. Why should a capacity to handle ordinary counterfactuals confer a capacity to handle such peculiar ones too?

Reply: That is like asking why a capacity to handle inferences between complex empirical sentences should confer a capacity to handle inferences involving logical truths and falsehoods too. There is no easy way to have the former without the latter. More specifically, developing a counterfactual supposition includes reasoning from it, and we cannot always tell in advance when such reasoning will yield a contradiction (there are surprises in logic). The undecidability of logical truth for first-order logic implies that there is no total mechanical test for the consistency of first-order sentences. Thus the inconsistent ones cannot be sieved out in advance (consider "In the next village there is a barber who shaves all and only those in that village who do not shave themselves"). Consequently, a general capacity to develop counterfactual suppositions must confer in particular the capacity to develop those which subsequently turn out inconsistent. Although the capacity may not be of uniform reliability, as already noted, the variation is primarily with the *antecedent* of the counterfactual (the supposition under development), not with its consequent (which is what is exceptional in (17) and (18)). In deductive inference, our reasoning to contradictions (as in proof by *reduction ad absurdum*) is not strikingly more or less reliable than the rest of our deductive reasoning.

Objection: The assumption about vacuous truth on which the account relies is wrong (Nolan 1997). For some counterpossibles (counterfactuals with metaphysically impossible antecedents) are false, such as (28), uttered by someone who mistakenly believes that he answered "13" to "What is 5 + 7?"; in fact he answered "11":

(28) If 5 + 7 were 13 I would have got that sum right.

Thus, contrary to (17), $\Box \mathbf{A}$ may be true while $\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \perp$ is false. In the argument for (17) in § 3, the objectionable premise is NECESSITY. If some worlds are metaphysically impossible, and $\mathbf{A}$ is true at some of them but false at all metaphysically possible worlds, while $\mathbf{B}$ is false at all worlds whatsoever, then every metaphysically possible $\mathbf{A}$ world is a $\mathbf{B}$ world, but the closest $\mathbf{A}$ worlds are not $\mathbf{B}$ worlds.[10] Similar objections apply to the other purported equivalences (18)–(22).

Reply: If *all* counterpossibles were false, $\Diamond\mathbf{A}$ would be equivalent to $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{A}$, for the latter would still be true whenever $\mathbf{A}$ was possible; correspondingly, $\Box\mathbf{A}$ would be equivalent to the dual $\neg(\neg\mathbf{A} \mathrel{\Box\!\!\rightarrow} \neg\mathbf{A})$ and one could carry out the programme of § 3 using the new equivalences. But that is presumably not what the objector has in mind. Rather, the idea is that the truth-value of a counterpossible can depend on its consequent, so that (28) is false while (29) is true:

(29) If $5 + 7$ were 13 I would have got that sum wrong.

However, such examples are quite unpersuasive.

First, they tend to fall apart when thought through. For example, if $5 + 7$ were 13 then $5 + 6$ would be 12, and so (by another eleven steps) 0 would be 1, so if the number of right answers I gave were 0, the number of right answers I gave would be 1.

Second, there are general reasons to doubt the supposed intuitions on which such examples rely. We are used to working with possible antecedents, and given the possibility of $\mathbf{A}$, the incompatibility of $\mathbf{B}$ and $\mathbf{C}$ implies that $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$ and $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{C}$ cannot both be true. Thus by over-projecting from familiar cases we may take the uncontentious (29) to be incompatible with (28). The logically unsophisticated make analogous errors in quantificational reasoning. Given the evident truth of "Every golden mountain is a mountain", they think that "Every golden mountain is a valley" is false, neglecting the case of vacuous truth. Since the logic and semantics of counterfactual conditionals is much less well understood, even the logically sophisticated may find similar errors tempting. Such errors may be compounded by a

[10] Technically, NECESSITY fails on a semanantics with similarity spheres for $\Box\!\!\rightarrow$ that include some impossible worlds (inaccessible with respect to $\Box$). Conversely, POSSIBILITY fails on a semantics with some possible worlds excluded from all similarity spheres (see Lewis 1986: 16 on universality). Inaccessible worlds seem not to threaten POSSIBILITY. For suppose that an $\mathbf{A}$ world w but no $\mathbf{B}$ world is accessible from a world v. Then if $\mathbf{A} \mathrel{\Box\!\!\rightarrow} \mathbf{B}$ holds at v on the usual semantics, there is an $\mathbf{A}$ world x such that every $\mathbf{A}$ world as close as x is to v is a $\mathbf{B}$ world. It follows that w is not as close as x is to v and that x is inaccessible from v, which contradicts the plausible assumption that any accessible world is at least as close as any inaccessible world.

tendency to confuse negating a counterfactual conditional with negating its consequent, given the artificiality of the constructions needed to negate the whole conditional unambiguously ("it is not the case that if ..."). Thus the truth of $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{B}$ (with **A** impossible) may be mistaken for the truth of $\neg(\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{B})$ and therefore the falsity of $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{B}$.

Some objectors try to bolster their case by giving examples of mathematicians reasoning from an impossible supposition **A** ("There are only finitely many prime numbers") in order to reduce it to absurdity. Such arguments can be formulated using a counterfactual conditional, although they need not be. Certainly there will be points in the argument at which it is legitimate to assert $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{C}$ (in particular, $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \mathbf{A}$) but illegitimate to assert $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{C}$ (in particular, $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{A}$). But of course that does not show that $\mathbf{A} \mathbin{\Box\!\!\rightarrow} \neg\mathbf{A}$ is false. At any point in a mathematical argument there are infinitely many truths that it is not legitimate to assert, because they have not yet been proved (Lewis 1986: 24–6 pragmatically explains away some purported examples of false counterfactuals with impossible antecedents).

We may also wonder what logic of counterfactuals the objectors envisage. If they reject elementary principles of the pure logic of counterfactual conditionals, that is an unattractive feature of their position. If they accept all those principles, then they are committed to operators characterized as in (17) and (18) that exhibit all the logical behaviour standardly expected of necessity and possibility. What is that modality, if not metaphysical modality?

A final problem for the objection is this. Here is a paradigm of the kind of counterpossible which the objector regards as false:

(30) If Hesperus had not been Phosphorus, Phosphorus would not have been Phosphorus.

Since Hesperus is Phosphorus, it is metaphysically impossible that Hesperus is not Phosphorus, by the necessity of identity. Nevertheless, the objectors are likely to insist that in imaginatively developing the counterfactual supposition that Hesperus is not Phosphorus, we are committed to the explicit denial of no logical truth, as in the consequent of (30). According to them, if we do our best for the antecedent, we can develop it into a logically coherent though metaphysically impossible scenario: it will exclude "Phosphorus is not Phosphorus". But they will presumably accept this trivial instance of reflexivity:

(31) If Hesperus had not been Phosphorus, Hesperus would not have been Phosphorus.

In general, however, coreferential proper names are intersubstitutable in counterfactual contexts. For example, the argument from (32) and (33) to (34) is unproblematically valid:

(32) If the rocket had continued on that course, it would have hit Hesperus.

(33) Hesperus = Phosphorus.

(34) If the rocket had continued on that course, it would have hit Phosphorus.

Similarly, the argument from (31) and (33) to (30) should be valid. But (31) and (33) are uncontentiously true. If the objector concedes that (30) is true after all, then there should be an explanation of the felt resistance to it, compatible with its truth, and we may reasonably expect that explanation to generalize to other purported examples of false counterpossibles. On the other hand, if objectors reject (30), they must deny the validity of the argument from (31) and (33) to (30). Thus they are committed to the claim that counterfactual conditionals create opaque contexts for proper names (the same argument could be given for other singular terms, such as demonstratives). But that is highly implausible. (32) and (34) are materially equivalent because their antecedents and consequents concern the same objects, properties and relations: it matters not that different names are used, because the counterfactuals are not about such representational features. But then exactly the same applies to (30) and (31). Their antecedents and consequents too concern the same objects, properties and relations. That the antecedent of (30) and (31) is in fact metaphysically impossible does not radically alter their subject matter. The transparency of the counterfactual conditional construction concerns its general logical form, not the specific content of the antecedent. Under scrutiny, the case for false counterpossibles looks feeble.

Objection: Counterfactuals are desperately vague and context-sensitive; equivalences such as (17) and (18) will infect □ and ◊, interpreted as metaphysical modalities, with all that vagueness and context-sensitivity.

Reply: Infection is not automatic. For instance, within a Lewis-Stalnaker framework, different readings or sharpenings of $\Box\!\!\rightarrow$ may differ on the similarity ordering of worlds while still agreeing on what worlds there are, so that the differences cancel out in the right-hand sides of (17) and (18). Whether a given supposition counterfactually implies a contradiction may be unclear to us; that does not imply that there is no right answer.

Objection: It has been argued that counterfactual conditionals lack truth-values (Edgington 2003, Bennett 2003: 252–6). If so, the assimilation of claims of metaphysical possibility and necessity to counterfactuals will deprive such claims of truth-values.

Reply: The issues are too complex to discuss properly here, but the readily intelligible occurrence of counterfactual conditionals embedded in the scope of other operators as in (23) and (24) is hard to make sense of without attributing truth-values to the embedded occurrences. Here is another example:

(35) Every field that would have been flooded if the dam had burst was ploughed.

(35) can itself be intelligibly embedded in more complex sentences in all the usual ways. In order to understand how such embeddings work, we must assign truth-conditions to (35); *ad hoc* treatments of a few particular embeddings are not enough. For (35) to have truth-conditions, "field that would have been flooded if the dam had burst" must have application-conditions. Thus there must be a distinction between the fields to which "would have been flooded if the dam had burst" applies and those to which it does not. But that is just to say that there must be a distinction between the values of "*x*" for which "If the dam had burst, *x* would have been flooded" is true and those for which it is false. That it is somewhat obscure what the truth-conditions of counterfactual conditionals are, and that we sometimes make conflicting judgments about them, hardly shows that they do not exist.

§ 5.

The counterfactual conditional is of course not the only construction in ordinary use that is closely related to metaphysical modality. Consider comments after a swiftly extinguished fire in an explosives factory:

(36) There could have been a huge explosion.

(37) There could easily have been a huge explosion.

The truth-value of both (36) (so interpreted) and (37) depends on the location of the fire, the precautions in place, and so on. The mere metaphysical possibility of a huge explosion is insufficient to verify either (36) (so interpreted) or (37). The restricted nature of the possibility is explicit in (37) with the word "easily"; it is implicit in the context of (36).[11] To discover the truth-value of (36) or (37), we need background information. We may also need our imagination, in attempting to develop a feasible scenario in which there is a huge explosion. We use the same general cognitive faculties as we do in evaluating related counterfactual conditionals, such as (38):

(38) If the fire engine had arrived a minute later, there would have been a huge explosion.

Judgments of limited possibility such as (36) (interpreted as above) and (37) have a cognitive value for us similar to that of counterfactual conditionals such as (38).

Both (36) and (37) entail (39), although not vice versa:

(39) It is metaphysically possible that there was a huge explosion.

This is another way in which our ordinary cognitive capacities enable us to recognize that something non-actual is nevertheless metaphysically possible. But we cannot reason from the negation of (36) or of (37) to the negation of (39).

Can metaphysical possibility be understood as the limiting case of such more restricted forms of possibility? Perhaps, but we would need some account of what demarcates the relevant forms of possibility from irrelevant ones, such as epistemic possibility. It also needs to be explained how, from the starting-point of ordinary thought, we manage to single out

[11] On easy possibility see Sainsbury 1997, Peacocke 1999: 310–28 and Williamson 2000: 123–30. On the idea that natural language modals such as "can" and "must" advert to contextually restricted ranges of possibilities see Kratzer 1977.

the limiting case, metaphysical modality. The advantage of counterfactual conditionals is that they allow us to single out the limiting case simply by putting a contradiction in the consequent; contradictions can be formed in any language with conjunction and negation Anyway, the connections with restricted possibility and with counterfactual conditionals are not mutually exclusive, for they are not being interpreted as rival semantic analyses, but rather as different cases in which the cognitive mechanisms needed for one already provide for the other.

The epistemology of metaphysical modality requires no dedicated faculty of intuition. It is simply a special case of the epistemology of counterfactual thinking, a kind of thinking tightly integrated with our thinking about the spatio-temporal world. To deny that such thinking ever yields knowledge is to fall into an extravagant scepticism. Here as elsewhere, we can do philosophy on the basis of general cognitive capacities that are in no deep way peculiarly philosophical.[12]

REFERENCES

Anderson, A. R. (1951). 'A note on subjunctive and counterfactual conditionals', *Analysis* 12, 35–8.

Bennett, J. (2003). *A Philosophical Guide to Conditionals*, Clarendon Press, Oxford.

Blackburn, S. (1987). 'Morals and modals', *Fact, Science and Morality*, (eds.) G. Macdonald and C. Wright, Blackwell, Oxford.

[12] An earlier version of this paper was presented at the 2005 Erfurt conference on philosophical knowledge. It developed out of Williamson 2005, especially 15–22, from which it inherits numerous debts; Williamson 2004 explains other aspects of the associated general account of philosophical knowledge. Further debts were acquired from discussion after the presentation of that paper as a Presidential Address to the Aristotelian Society and of related material in the Blackwell Brown lectures at Brown University, the Anders Wedberg lectures at Stockholm University (where Sören Häggqvist and Anna-Sara Malmgren commented on the relevant lectures), at the third meeting of the Portugese Society for Analytic Philosophy in Lisbon and at conferences and workshops at the Centre for Advanced Studies in the Norwegian Academy of Sciences in Oslo, the Australian National University, Rutgers University and the universities of Bristol, Munich and Rochester, at colloquia at the University of California Los Angeles and the universities of Arizona, Bologna, Heidelberg, Leeds, Nottingham, Turin and Warwick, and in Oxford, although the ideas presented here were not always centre stage. Thanks to all the individuals who have helped improve this paper, and in particular to Ann-Sara Malmgren and Thomas Kroedel for many valuable discussions of its themes.

Byrne, R. M. J. (2005). *The Rational Imagination: How People Create Alternatives to Reality*, MIT Press, Cambridge, Mass.

Collins, J., Hall, N. and Paul, L. A. (2004). *Causation and Counterfactuals*, MIT Press, Cambridge, Mass.

Craig, E. (1985). 'Arithmetic and fact', *Essays in Analysis*, (ed.) I. Hacking, Cambridge University Press, Cambridge.

Currie, G. (1995). 'Visual imagery as the simulation of vision', *Mind and Language* 10, 17–44.

Davies, M. and Stone, T. (eds.) (1995). *Mental Simulation: Evaluations and Applications*, Blackwell, Oxford.

Edgington, D. (2003). 'Counterfactuals and the benefit of hindsight', *Causation and Counterfactuals*, (eds.) P. Dowe and P. Noordhof, Routledge, London.

Evans, J. St. B. T. and Over, D. E. (2004). *If*, Oxford University Press, Oxford.

Gendler, T. Szabó and Hawthorne, J. (eds.) (2002). *Conceivability and Possibility*, Clarendon Press, Oxford.

Goldman, A. (1992). 'Empathy, mind and morals', *Proceedings and Addresses of the American Philosophical Association* 66/3, 17–41.

Goodman, N. (1955). *Fact, Fiction, and Forecast*, Harvard University Press, Cambridge, MA.

Harris, P. (2000). *The Work of the Imagination*, Blackwell, Oxford.

Hill, C. S. (2006). 'Modality, modal epistemology, and the metaphysics of consciousness', *The Architecture of the Imagination: New Essays on Pretense, Possibility and Fiction*, (ed.) S. Nichols, Oxford University Press, Oxford.

Jackson, F. (1977). 'A causal theory of counterfactuals', *Australasian Journal of Philosophy* 55, 3–21.

Kahneman, D. and Tversky, A. (1982). 'The simulation heuristic', *Judgement under Uncertainty*, (eds.) D. Kahneman, P. Slovic and A. Tversky, Cambridge University Press, Cambridge.

Kaplan, D. (1989). 'Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals', *Themes from Kaplan*, (eds.) J. Almog, J. Perry and H. Wettstein, Oxford University Press, Oxford.

Kratzer, A. (1977). 'What "must" and "can" must and can mean', *Linguistics and Philosophy* 1, 337–355.

Kripke, S. A. (1980). *Naming and Necessity*, Blackwell, Oxford.

Lewis, D. (1973a). 'Counterfactuals and comparative possibility', *Journal of Philosophical Logic* 2, 418–46. Reprinted in his *Philosophical Papers*, vol. 2, Oxford University Press, Oxford, 1986, to which page numbers refer.

— (1973b). 'Causation', *Journal of Philosophy* 70, 556–67.

— (1979). 'Counterfactual dependence and time's arrow', *Noûs* 13, 455–476.

— (1986). *Counterfactuals*, revised edn. Harvard University Press, Cambridge, Mass.

McKinsey, M. (1991). 'Anti-individualism and privileged access', *Analysis* 51, 9–16.

Nichols, S. and Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding of Other Minds*, Clarendon Press, Oxford.

Nichols, S., Stich, S. P., Leslie, A. and Klein, D. (1996). 'Varieties of off-line simulation', *Theories of Theories of Mind*, (eds.) P. Carruthers and P. K. Smith, Cambridge University Press, Cambridge.

Nolan, D. (1997). 'Impossible worlds: a modest approach', *Notre Dame Journal for Formal Logic* 38, 535–572.

Nozick, R. (1981). *Philosophical Explanations*, Clarendon Press, Oxford.

Peacocke, C. (1985). 'Imagination, experience and possibility', *Essays on Berkeley: A Tercentennial Celebration*, (eds.) J. Foster and H. Robinson, Clarendon Press, Oxford.

— (1999). *Being Known*, Clarendon Press, Oxford.

Ramsey, F. P. (1978). *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*, (ed.) D. H. Mellor. Routledge & Kegan Paul, London.

Roese, N. J. and Olson, J. (1993). 'The structure of counterfactual thought', *Personality and Social Psychology Bulletin* 19, 312–19.

— (1995). 'Functions of counterfactual thinking', *What Might Have Been: The Social Psychology of Counterfactual Thinking*, (eds.) N. J. Roese and J. M. Olson, Erlbaum, Mahwah, NJ.

Sainsbury, R. M. (1997). 'Easy possibilities', *Philosophy and Phenomenological Research*, 57, 907–19.

Schechter, J. (2006). 'Can evolution explain the reliability of our logical beliefs', typescript.

Shope, R. K. (1983). *The Analysis of Knowing: A Decade of Research*, University Press Princeton, Princeton.

Stalnaker, R. (1968). 'A theory of conditionals', *American Philosophical Quarterly Monographs* 2 (*Studies in Logical Theory*). 98–112.

— (1999). *Context and Content*, Oxford University Press, Oxford.

Williams, B. (1966). 'Imagination and the self', *Proceedings of the British Academy* 52, 105–124.

Williamson, T. (2000). *Knowledge and its Limits*, Oxford University Press, Oxford.

— (2004). 'Philosophical "intuitions" and skepticism about judgement', *Dialectica* 58, 109–153.

— (2005). 'Armchair philosophy, metaphysical modality and counterfactual thinking', *Proceedings of the Aristotelian Society* 105, 1–23.

— (2006). ‘Indicative versus subjunctive conditionals, congruential versus non-hyperintensional contexts’, *Philosophical Issues* 16, 310–33.

Wright, C. (1989). ‘Necessity, caution and scepticism’, *Aristotelian Society* sup. 63, 203–38.

Grazer Philosophische Studien
74 (2007), 125–141.

THE POSSIBILITY OF KNOWLEDGE

Quassim CASSAM
University College London

Summary

I focus on two questions: what is knowledge, and how is knowledge possible? The latter is an example of a how-possible question. I argue that how-possible questions are obstacle-dependent and that they need to be dealt with at three different levels, the level of means, of obstacle-removal, and of enabling conditions. At the first of these levels the possibility of knowledge is accounted for by identifying means of knowing, and I argue that the identification of such means also contributes to a proper understanding of what knowledge is.

1. *Introduction*

I'm going to be addressing two questions here. The first, which I will call the "what" question is: what is knowledge? The second, which I will call the "how" question is: how is knowledge possible?[1] As well as attempting to give answers to these questions I want to say something about the relationship between them and the proper methodology for answering them. By "knowledge" I mean propositional knowledge, the knowledge that something is the case. I am going to suggest that the standard approaches to the "what" and "how" questions are defective and that the key to answering both questions is the notion of a *means* of knowing. In brief, my idea is that the way to explain how knowledge is possible is to identify various means by which it is possible and that the identification of the means by which knowledge is possible contributes to a proper understanding of what knowledge fundamentally *is*.

To bring my proposal into focus, I would like to start by outlining some contrasting approaches. One standard approach to the "what" question

[1] The "what" and "how" questions are two of the three questions which Hilary Kornblith describes as being among the central questions of epistemology. The third question is "What should we do in order to attain knowledge?" (Kornblith 1999: 159).

is the *analytic* approach. This suggests that to ask what knowledge is is to ask what it is to know that something is the case.[2] This is taken to be a question about the truth conditions rather than the meaning of statements of the form "S knows that p".[3] Suppose, for example, that I know that the cup into which I'm pouring coffee is chipped. The analytic approach says that a good account of what it is to know this will be an account of the necessary and sufficient conditions for knowing that the cup is chipped, and that the proper methodology for identifying these conditions is conceptual analysis, conceived of as a form of armchair philosophical reflection. The idea is that by analysing the concept of knowledge into more basic concepts one discovers necessary and sufficient conditions for knowing that the cup is chipped and thereby explains what it is to know that something is the case.

The familiar problem with this approach is that it is actually very difficult to come up with necessary and sufficient conditions for propositional knowledge that are both non-circular and correct.[4] As Williamson points out, there seem to be counterexamples to every existing analysis and it's not clear in any case that a complicated analysis that somehow managed not to succumb to the usual counterexamples would necessarily tell us very much about knowledge is. But if we conclude on this basis that the pursuit of analyses is "a degenerating research programme" (Williamson 2000: 31) then analytic epistemology leaves us without an answer to the "what" question.

One reaction to these difficulties has been to argue that the fundamental mistake of analytic epistemology is that it focuses on the *concept* of knowledge rather than on *knowledge*. According to Kornblith, for

[2] This way of thinking about the "what" question is suggested by Alvin Goldman. See Goldman 1986: 42.

[3] Goldman emphasizes the distinction giving the meaning and giving the truth conditions of "S knows that p" in the concluding paragraphs of 'A Causal Theory of Knowing', originally published in 1967 and reprinted in Goldman 1992.

[4] Gettier 1963 provides an early illustration of some of these difficulties. Gettier shows that the traditional justified-true-belief analysis of knowledge is incorrect because truth, belief and justification aren't sufficient for knowledge. Gettier-style counterexamples to the traditional analysis can be dealt with by beefing up the notion of justification but this threatens circularity. As Williamson points out, "if someone insists that knowledge *is* justified true belief on an understanding of 'justified true belief' strong enough to exclude Gettier cases but weak enough to include ordinary empirical knowledge, the problem is likely to be that no standard of justification is supplied independent of knowledge itself" (2000: 4). This is only a problem for those analytic epistemologists who are looking for a reductive definition of knowledge in terms of more basic concepts.

example, "the subject matter of epistemology is knowledge itself, not our concept of knowledge" (Kornblith 2002: 1) and "knowledge itself" is a natural kind. This implies that we should go in for a *naturalistic* rather than an analytic approach to the "what" question. Specifically, the proposal is that if knowledge is a natural kind then we should expect work in the empirical sciences rather than armchair conceptual analysis to be the key to understanding what it is. But knowledge isn't a natural kind. There are too many disanalogies between knowledge and genuine natural kinds for this to be plausible, and in practice those who try to "naturalize" epistemology either end up ignoring the what question altogether or answering it on the basis of just the kind of armchair reflection that analytic epistemologists go in for.[5]

If this isn't bad enough, the "how" question seems no less intractable. One worry is that we can't explain how knowledge is possible if we don't know what knowledge is, so if we can't answer the "what" question then we can't answer the "how" question either. The standard approach to the "how" question is the *transcendental* approach, according to which the way to explain how knowledge is possible is to identify necessary conditions for its possibility. Yet it is hard to see how this helps. We can see what the

[5] Quine is someone in the naturalistic tradition who effectively ignores the "what" question. See Quine 1969. In contrast, Kornblith doesn't ignore it. He claims that knowledge requires reliably produced true belief and that he doesn't arrive at this conclusion by analysing the concept of knowledge. Yet in claiming that "knowledge is, surely, more than just true belief" (2002: 54) he seems to be relying on some form of armchair reflection; at any rate, it is hard to see how it can be an empirical question whether knowledge is or is not more than just true belief. As for the emphasis on reliability, this is Kornblith's explanation: "If we are to explain why it is that plovers are able to protect their nests, we must appeal to a capacity to recognize features of the environment, and thus the true beliefs that particular plovers acquire will be the product of a stable capacity for the production of true beliefs. The resulting true beliefs are not merely accidentally true; they are produced by a cognitive capacity that is attuned to its environment. In a word, the beliefs are reliably produced. The concept of knowledge which is of interest to us here thus requires reliably produced true belief" (2002: 58). What is obscure about this passage is the transition from the penultimate sentence to the last sentence. There might be empirical grounds for attributing reliably produced true beliefs to plovers but the further question is whether reliably produced true beliefs constitute knowledge. Kornblith doesn't explain how this can be established on empirical grounds. If belief, truth and reliability are sufficient for knowledge then attributions of reliably produced true beliefs to plovers are, *de facto*, attributions of knowledge but what, apart from armchair reflection, can tell us that belief, truth and reliability *are* sufficient for knowledge? Kornblith doesn't say. On the underlying issue of whether knowledge is a natural kind, knowledge doesn't have anything recognizable as a real essence in the way that natural kinds like gold and water have real essences. For Kornblith, however, natural kinds are "homeostatically clustered properties" (2002: 61) and this is the basis of his identification of knowledge as a natural kind. I don't have the space to go into this here.

problem is by thinking about scepticism. Sceptics ask how knowledge of the external world is possible given that we can't be sure that various sceptical possibilities do not obtain. It is not an answer to this question simply to draw attention to what is *necessary* for the existence of the kind of knowledge which the sceptic thinks we can't possibly have.[6] For example, it might be true that knowledge requires a knower but this observation leaves us none the wiser as to how knowledge of the external world is possible.

Let's agree, then, that we still don't have satisfactory answers to my two questions. So where do we go from here? We could try defending one or other of the standard approaches against the objections I have been discussing but this is not what I want to do here. As I have already indicated, I believe that a different approach is needed so now would be a good time to spell out what I have in mind. One of the features of my alternative is that it addresses the "how" question first and then moves on to the "what" question. The significance of doing things in this order should become clearer as I go along. In the meantime, let's start by taking a closer look at the "how" question, and about what is needed to answer it.

2. *How is knowledge possible?*

The first thing to notice is that what I have been calling "how" questions are really "how-possible" questions. This is worth pointing out because there are how questions that aren't how-possible questions.[7] Think about the difference between asking how John Major became Prime Minister in 1990 and asking how it was possible for John Major to become Prime Minister in 1990. To ask how Major became Prime Minister is to ask for an account of the stages or steps by which he became Prime Minister.[8] There is no implication that it is in any way surprising that he became Prime Minister or that there was anything that might have been expected to prevent him from becoming Prime Minister. There is such an implication when one asks how it was possible for Major to become Prime Minister.

[6] This needs to be qualified. Drawing attention to what is necessary for knowledge of the external world might help to defuse scepticism if it can be shown that the necessary conditions do *not* include the knowledge that the sceptic's possibilities don't obtain. This is what I refer to below as an *obstacle-dissipating* response to scepticism. The fact remains, however, that necessary conditions *per se* are not to the point.

[7] William Dray makes this point in Dray 1957: 166. My account of how-possible questions is much indebted to Dray's valuable discussion.

[8] Cf. Dray 1957: 166.

The implication is that there was some obstacle to such a thing happening, and this is what gives the how-possible question its point. For example, one might think that Major's social and educational background ought to have made it impossible for him to become Prime Minister.[9] The fact is, however, that he did become Prime Minister. So what one wants to know is not *whether* it happened, because it did, but *how it could have* happened, how it was possible.

On this account, how-possible questions are *obstacle-dependent* in a way that simple how questions are not.[10] One asks how X is possible on the assumption that there is an obstacle to the existence or occurrence of X. What one wants to know is how X is possible despite the obstacle. The most striking how-possible questions are ones in which the obstacle looks like making the existence or occurrence of X not just surprising or difficult but impossible. In such cases the challenge is to explain how something which looks impossible is nevertheless possible. One way of doing this would be to show that the obstacle which was thought to make X impossible isn't genuine. This would be an *obstacle-dissipating* response to a how-possible question. In effect, this response rebuts the presumption that X isn't possible and thereby deprives the how-possible question of its initial force. Another possibility would be to accept that the obstacle is genuine and to then explain how it can be overcome. This would be an *obstacle-overcoming* response to a how-possible question.

We can illustrate the distinction between dissipating and overcoming an obstacle by turning from British politics to Prussian epistemology and looking at one of Kant's many how-possible questions in the first *Critique*. The question is: how is mathematical knowledge possible? What gives this question its bite is the worry that mathematical knowledge can't be accounted for by reference to certain presupposed basic sources of knowledge. The two presupposed sources are experience and conceptual analysis. Assuming that mathematical truths are necessarily true our knowledge of them can't come from experience; it must be *a priori* knowledge because experience can only tell us that something is so not that it must be so. Assuming that mathematical truths are synthetic it follows that conceptual analysis can't be the source of our knowledge of them either. So if

[9] Unlike most modern British Prime Ministers Major didn't attend university. His father was a trapeze artist.

[10] See Dray 1957: 156–69 for a defence of this conception of how-possible questions. Dray's ideas have also been taken up by Robert Nozick and Barry Stroud. See Nozick 1981: 8–10, and Stroud 1984: 144.

experience and conceptual analysis are our only sources of knowledge then mathematical knowledge is impossible. Let's call this apparent obstacle to the existence of mathematical knowledge the *problem of sources*. It is this problem which leads Kant to ask *how* mathematical knowledge is possible because he doesn't doubt that synthetic *a priori* mathematical knowledge *is* possible.

An obstacle-dissipating response to Kant's question would dispute the assumption that neither experience nor conceptual analysis can account for our mathematical knowledge. For example, conceptual analysis can account for it if mathematical truths are analytic rather than synthetic. Alternatively, there is no reason why mathematical knowledge couldn't come from experience if the truths of mathematics aren't necessary or if it is false that experience can't tell us that something must be so. Each of these dissipationist responses to Kant's question amounts to what might be called a *presupposed sources* solution to the problem of sources; in each case the possibility of mathematical knowledge is accounted for by reference to one of the presupposed sources of knowledge. But this isn't Kant's own preferred solution. His solution is an *additional sources* solution since it involves the positing of what he calls "construction in pure intuition" as an additional source of knowledge by reference to which at least the possibility of geometrical knowledge be accounted for.[11] This is an obstacle-overcoming rather than an obstacle-dissipating response to a how-possible question because it doesn't dispute the existence of the obstacle which led the question to be asked in the first place; it accepts that the obstacle is, in a way, perfectly genuine and tries to find a way around it.[12]

The only sense in which construction in pure intuition, the use of mental diagrams in geometrical proofs, is an "additional" source of knowledge is that no account was taken of it in the discussion leading up to the raising of the how-possible question. It isn't additional in the sense that geometers haven't been using it all along. By identifying construction in intuition as a means of acquiring synthetic *a priori* geometrical knowledge

[11] Kant describes the role of construction in geometrical proof in the chapter of the first *Critique* called 'The Discipline of Pure Reason in its Dogmatic Employment'. See, especially, A713/ B741. References in this form are to Kant 1932.

[12] Clearly, the only sense in which Kant accepts that the obstacle is genuine is that mathematical knowledge can't be accounted for *if* experience and conceptual analysis are its only possible sources. In another sense he doesn't think that the obstacle is genuine because he thinks that it is false that experience and conceptual analysis are the only possible sources of mathematical knowledge. This suggests that the distinction between overcoming and dissipating an obstacle isn't always a sharp one and that overcoming an obstacle can shade off into obstacle-dissipation.

Kant explains how such knowledge is possible. In general, drawing attention to the means by which something is possible is a means of explaining how it is possible yet the means by which something is possible needn't be necessary conditions for its possibility. Catching the Eurostar is a means of getting from London to Paris in less than three hours but not a necessary condition for doing this. So if all one needs in order to explain how something is possible is to identify means by which it is possible then there is no need to look for necessary conditions.

But is it plausible that the identification of means of knowing suffices to explain how knowledge is possible? Not if it is unclear how one can acquire the knowledge that is in question by the proposed means. For example, one worry about Kant's account of geometry is that what is constructed in intuition is always a specific figure whereas the results of construction are supposed to be universally valid propositions. How then, is it possible for construction to deliver knowledge of such propositions? According to Kant there is no problem as long as constructed figures are determined by certain rules of construction which he calls "schemata". As he puts it, the single figure which we draw serves to "express" the concept of a triangle because it is "determined by certain universal conditions of construction".[13]

For present purposes the details of account are much less interesting than its structure. What we can extract from Kant's discussion is the suggestion that his how-possible question needs to be dealt with at a number of different levels. First there is the level of *means*, the level at which the possibility of mathematical knowledge is accounted for by identifying means by which it is possible. Second, there is the level of *obstacle-removal*, the level at which obstacles to the acquisition of mathematical knowledge by the proposed means are overcome or dissipated. But this still isn't the end of Kant's story. He thinks that even after the problem of accounting for the universality of mathematical knowledge has been solved there is a further question that naturally arises. This further question is: *what makes it possible* for construction in intuition to occur and to be a source of mathematical knowledge?

This last question concerns the background necessary conditions for the acquisition of mathematical knowledge by constructing figures in intuition. What it seeks is not a way round some specific obstacle but, as it were, a positive explanation of the possibility of acquiring a certain kind

[13] A714/ B742.

of knowledge by certain specified means. We have now reached what can be called the level of *enabling conditions*.[14] Kant's proposal at this level is that what makes it possible for mental diagrams to deliver knowledge of the geometry of physical space is the fact that physical space is subjective.[15] If space were a "real existence" in the Newtonian sense it wouldn't be intelligible that intuitive constructions are capable of delivering knowledge of its geometry. That is why, according to Kant, we must be transcendental idealists if we want to understand how geometrical knowledge is possible. So this looks like a third explanatory level in addition to the level of means and that of obstacle-removal.

In fact, the distinction between the second and third levels isn't a sharp one in this case. If space were a real existence then that would be an obstacle to the acquisition of geometrical knowledge by means of construction. This makes it appear that what happens at the level of enabling conditions is much as exercise in obstacle-removal as what happens at the second level. Yet there are other how-possible questions in connection with which there is a sharper distinction between the second and third levels, and I now want to examine one such question. In any case, we shouldn't be reading too much into Kant's account of geometry because it isn't as if we still think about geometry in the way that he thought about it. In particular, if geometrical knowledge isn't synthetic *a priori* then we don't have Kant's reasons for worrying about how it is possible. But I now want to show that the basic framework of his discussion can be used to think about a range of different how-possible questions.

As we have seen, sceptics ask how knowledge of the external world is possible given that we can't be sure that various sceptical possibilities do not obtain. Take an ordinary proposition about the external world such as the proposition that the cup into which I am pouring coffee is chipped. How is it possible for me to know that this is the case? The obvious answer would be: by seeing that it is chipped, or feeling that it is chipped, being told by the person sitting opposite me that it is chipped, and so on. Seeing that the cup is chipped, which is a form of what Dretske calls

[14] For more on the notion of an enabling condition see Dretske 1969: 82–3. Dretskean enabling conditions are empirical whereas Kantian enabling conditions are *a priori*. An empirical enabling condition is one which can only be discovered by empirical investigation. An *a priori* condition can be discovered without any empirical investigation.

[15] Subjective in the transcendental idealist sense, according to which space belongs "only to the form of intuition" (A23/ B38). This is supposed to be compatible with space's being "empirically real".

'epistemic seeing', looks like a means of knowing that it is chipped.[16] But now we come up against the sceptic's obstacle. The sceptic thinks that I can't correctly be said to see that the cup is chipped unless I can eliminate the possibility that I am dreaming, and that I can't possibly eliminate this possibility.[17] This is a version of the problem of sources. The obstacle to the acquisition of perceptual knowledge, to knowing that the cup is chipped by seeing that it is chipped, takes the form of an epistemological requirement that supposedly can't be met. In fact, it is the precisely the obstacle that might have prompted one to ask the how-possible question in the first place.

As usual, we can either try to overcome the obstacle or dissipate it. To overcome the obstacle would be to show that it is possible to eliminate the possibility that one is dreaming.[18] To dissipate the obstacle would be to show that there is no such epistemological requirement on epistemic seeing. This looks like the best bet. When one understands the sceptic's requirement in the way that he understands it one sees that one couldn't possibly meet it, and that is why the only hope of dealing with the apparent obstacle to knowing about the external world by means of the senses is to show that it isn't genuine. One way of doing this would be to argue that we are less certain of the correctness of the sceptic's obstacle-generating epistemological requirement than we are of the knowledge that it purports to undermine, for example the knowledge that the cup is chipped.[19] Epistemological requirements mustn't have unacceptable consequences, and it is an unacceptable consequence of the sceptic's requirement that it makes it impossible to know such things. To the extent that knowing that one isn't dreaming is a requirement on anything in this area it is a requirement on knowing that one sees that the cup is chipped, not a requirement on seeing that the cup is chipped.

It is controversial whether these attempts at obstacle-dissipation are successful but let's assume for present purposes that they are. So we now have the idea that epistemic seeing is a means of knowing about the external world, though obviously not the only means, together with the

[16] There is a detailed account of the notion of epistemic seeing in Dretske 1969. See, especially, Ch. 3.

[17] See Stroud 1984, especially Ch. 1, for more on this sceptical argument.

[18] See McDowell 1998: 238–9 for something along these lines though McDowell is careful not to claim that it is possible to meet the sceptic's requirement on the sceptic's own terms.

[19] This is a version of what Baldwin calls G. E. Moore's argument from "differential certainty". See Baldwin 1990: 269–74 and Moore 1953.

suggestion that there isn't a genuine obstacle to knowing about the external world by such means. This is a presupposed sources solution to the problem of sources since perception is a presupposed source of knowledge of the external world. We reach the level of enabling conditions when we ask what makes it possible to see that the cup is chipped and thereby to know that it is chipped. We don't have to ask this question but we can ask it.[20] Here, then, are two Kantian thoughts: in order to see that the cup is chipped I must be able to see the cup, and the cup itself is an object. To see an object I must be able to see some of its spatial properties so the enabling conditions for seeing that the cup is chipped include possession of a capacity for spatial perception. They also include a capacity for categorial thinking on the assumption that one couldn't see that the cup is chipped if one lacked the concept *cup* and that one couldn't have this concept if one lacked a repertoire of categorial concepts such as substance, unity, plurality and causality.[21]

The identification of these enabling conditions for epistemic seeing isn't an exercise in obstacle-removal in the way that Kant's account of the enabling conditions for the acquisition of geometrical knowledge is an exercise in obstacle-removal. It is true that when a necessary condition for a particular cognitive achievement isn't fulfilled the very fact that it isn't fulfilled *becomes* an obstacle to that achievement but it still doesn't follow that the point of talking about enabling conditions must be to deal with some pre-existing obstacle. For example, there is no such obstacle to seeing the cup is chipped that is dissipated or overcome by the observation that it wouldn't be possible to see such a thing without a capacity for spatial perception. The question, "what makes X possible?", is an *explanation-seeking* question, and there is more to explaining what makes X possible than showing that there is nothing that makes it impossible.

To sum up, we now have a *multi-levels* framework for dealing with how-possible questions in epistemology. When we find ourselves faced with a how-possible question which asks how knowledge of a certain kind is possible, we start by identifying means by which it is possible to acquire this kind of knowledge. This is what I have been calling the level of means.

[20] We don't have to ask it because it's not obvious that an explanation of the possibility of knowledge of the external world which doesn't talk about enabling conditions is incomplete.

[21] One could see a chipped cup without having the concept *cup* but seeing *that* the cup is chipped is a different matter. Williamson uses a different example to make the same point in Williamson 2000: 38. For a defence of the idea that empirical concepts presuppose categorial concepts see Longuenesse 1998.

Then we set about trying to remove obstacles to acquiring knowledge by the proposed means. This is the obstacle-removing level. Finally, we might ask what makes it possible to acquire knowledge by the suggested means and this takes us to the level of enabling conditions. So how does this way of approaching how-possible questions differ from the transcendental approach? The main difference is in the significance that the two approaches attach to necessary conditions. The transcendental approach tries to explain how knowledge is possible by reference to its necessary conditions and I have already explained why this isn't right. Nobody would think that explaining how it is possible to get from London to Paris in less than three hours is a matter of identifying necessary conditions for getting from London to Paris in less than three hours and it is no more plausible that explaining how it is possible to know that p, where p is a proposition about the external world or other minds or whatever, is matter of identifying necessary conditions for knowing that p. In both cases, means rather than necessary conditions are the first thing we should be looking for.

This is not to deny that necessary conditions have a part to play in a multiple levels framework. Enabling conditions are, after all, necessary conditions but this doesn't mean that a multiple levels explanation of the possibility of knowledge is a transcendental explanation. The necessary conditions which figure in transcendental explanations are universal in scope. For example, there is the suggestion that the perception of space is a necessary condition for the acquisition of *any* empirical knowledge, regardless of the specific means by which it is acquired.[22] Yet it seems unlikely that the role of spatial awareness in coming to know that p by hearing that p or reading that p will be anything like its role in coming to know that p by seeing that p. Nevertheless, seeing that p, hearing that p and reading that p are all ways of acquiring empirical knowledge. What this suggests is that the necessary conditions which figure in transcendental explanations are excessively general. The same isn't true of the necessary conditions which figure in multiple levels explanations because these conditions can be *means-specific*. There is no commitment in this framework to the idea that the background necessary conditions for knowing that p by seeing that p are bound to be the same as the background necessary conditions for knowing that p by hearing that p; they might be but needn't be.

If, as I have been claiming, the transcendental approach to explaining how knowledge is possible isn't the right one why has it been so popular?

[22] See Strawson 1997 for a suggestion along these lines.

One explanation is that showing *that* we know is sometimes confused with explaining *how* we know. So, for example, if we have experience, and knowledge of the external world is necessary for experience, then it follows that we have knowledge of the external world. But even if this transcendental argument is convincing on its own terms it doesn't explain *how* we know what it claims we do know; the thesis that knowledge of the external world is *necessary* doesn't explain how it is *possible* given all the obstacles that have been thought - mistakenly as it turns out - to make it impossible. That is why, if we are serious about explaining how knowledge is possible a different approach is needed, one which emphasizes means rather than necessary conditions.

3. *What is knowledge?*

With this discussion of how-possible questions in the background let us now turn to the "what" question. I want to defend the suggestion that that an effective way of explaining what knowledge is is to identify various means by which it is possible, and that the notion of a means of knowing therefore has as large a part to play in relation to the "what" question as in relation to how-possible questions. A good way of seeing the force of this suggestion would be to note that when we claim to know that something is the case there is a further question to which we are "directly exposed" (Austin 1979: 77). This further question is: how do you know? This is an example of a simple "how" question rather than a how-possible question and, as Austin points out, even simple how questions can be read in several different ways. For example, "how do you know that the cup is chipped?" can mean "how did you come to know that the cup is chipped?" or "how are you in a position to know that the cup is chipped?" or "how do *you* know that the cup is chipped?".

On the first of these three readings the simple "how" question is concerned with the *acquisition* of knowledge. Since there are lots of different ways of coming to know that the cup is chipped there are lots of different ways of answering the question.[23] Good answers to "how did you come to

[23] There is a mention of "ways of coming to know" in Stroud 2000. Stroud remarks that "there are countless ways of coming to know something about the world around us" (2000: 3) but that what we seek in philosophy isn't just a "list of sources". I am more sympathetic to the idea that an open-ended list of sources is precisely what we need if we want to understand "how we get the knowledge we have—to explain how it is possible" (ibid.).

know that the cup is chipped?" would include "by seeing that it is chipped" and "by feeling that it is chipped". A bad answer, in most circumstances, would be "by imagining that it is chipped". The important point, however, is that there must *be* an answer to the how-did-you-come-to-know question and that there is an intuitive distinction between good and bad answers to questions of this form. What we are reluctant to accept is that it can be a brute fact that someone knows without there being some specific way in which he came to know. It isn't possible to "just know" that the cup is chipped, and some ways of coming to know this are better than others.

This proposal is similar in some ways to Williamson's proposal that "if one knows that A then there is a specific way in which one knows" (2000: 34) but what I am calling "ways of coming to know" are different from Williamson's "ways of knowing". Ways of knowing are expressed in language by factive mental state operators (FMSOs).[24] Without going into too much detail, the basic idea is that if Φ is a FMSO then the inference from 'S Φs that p' to 'p' is deductively valid, as is the inference from 'S Φs that p' to 'S knows that p. In these terms, 'sees', 'regrets' and 'remembers' are all examples of FMSOs and are therefore also all examples of "ways of knowing" in Williamson's sense. In other words, if I see or regret or remember that the cup is chipped then the cup is chipped and I know that it is chipped. Yet only seeing that the cup is chipped is a way of *coming* to know that it is chipped, of *acquiring* this piece of knowledge; it would be distinctly odd to say that I came to know that the cup is chipped by regretting that it is chipped or even by remembering that it is chipped. Ways of coming to know are therefore sub-class of Williamson's "ways of knowing", and the present proposal is that what is needed to answer a how-did-you-come-to-know question is reference to a way of coming to know rather than to a mere "way of knowing".

How does this help with the "what" question? Suppose we agree that an account of what propositional knowledge is will need to be an account of what it is for a subject S to know that p. Having rejected the idea that explaining what it is for S to know that p is a matter of coming up with non-circular necessary and sufficient conditions for S to know that p we can now argue as follows: given that if S knows that p there must be some way in which S came to know that p, what it is for S to know that p can be understood by reference to the different ways in which it is possible for

[24] See Williamson 2000: 34–39 for more on the notion of a factive mental state operator.

someone like S to come to know something like p.[25] Since there might be countless ways of coming to know that p the notion of a way of coming to know that p is open-ended. The claim is that we get a fix on what it is to know that p by identifying good answers to the question "how do you know?" on the first of Austin's three readings of this question. In other words, we explain what propositional knowledge *is* by listing some of the ways of *acquiring* it; for example, we explain what it is to know that the cup is chipped by listing some of the ways of coming to know that the cup is chipped.

Ways of coming to know that p are means of knowing that p so we are now in a position to see why the notion of a means of knowing matters. Whether one is concerned with what it is to know that p, with how one knows that p, or with how it is possible to know that p it is difficult to exaggerate the importance of the notion of a means. Just as we have explained how it is possible to know that p by identifying means of knowing that p so we are now explaining what it is to know that p by identifying means of knowing that p. The identification of means of knowing that p is therefore a means explaining what it is to know that p just as it is a means of explaining how it is possible to know that p. So the position is not that one first tries to figure out what knowledge is and then tries to figure out how it can be acquired. Rather, one figures out what knowledge is *by* figuring out how it can be acquired.

To get a feel for this proposal consider the question "what is cricket?". An effective way of answering this "what" question would be to describe how cricket is played. Since one can learn what cricket is *by* learning how it is played it's no good objecting that one can't understand how cricket is played unless one already knows what it is. Similarly, it's no good objecting that one can't understand how knowledge is acquired unless one already knows what it is. Explaining what knowledge is by describing how it is acquired is like explaining what cricket is by describing how it is played.[26] In neither case is an answer to the "what" question presupposed and in neither case can the "what" question be answered by coming up with necessary and sufficient conditions. We wouldn't try to explain what cricket is

[25] This is not unlike Williamson's suggestion that "knowing that A is seeing that A or remembering or … that A, if the list is understood as open-ended, and the concept *knows* is not identified with the disjunctive concept" (2000: 34). There is much more on Williamson, and on the differences between his approach and mine, in Cassam, forthcoming.

[26] The analogy isn't perfect. There are lots of ways of acquiring knowledge but it isn't true in the same sense that there are lots of ways of playing cricket.

by specifying necessary and sufficient conditions for a game to be a game of cricket and we shouldn't try to explain what knowledge is by specifying necessary and sufficient conditions for a belief to constitute knowledge.

The "means" approach which I have been recommending might need supplementing in various ways. For example, knowledge can be retained and transmitted as well as acquired so a fuller picture of what knowledge is might need to say something about some of the different ways of retaining and transmitting it as well as some of the different ways of acquiring it. It might also need to be recognized that there are some things that we can't know because the obstacles to knowing them can't be overcome or dissipated. Perhaps some propositions about the distant past are like this. And even in the case of things that we are capable of knowing, some ways of coming to know them might be more basic than others. For example, seeing that the cup is chipped might count as in some sense a more basic way of coming to know that it is chipped than reading in a newspaper that it is chipped.

Finally, more needs to be said about the distinction between good and bad answers to how-did-you-come-to-know questions. A good answer to one such question might be a bad answer to another. For example, "by constructing a figure in pure intuition" might be a good answer to "how did you come to know that the internal angles of triangle are equal to two right angles?" but a bad answer to "how did you come to know that the cup is chipped?". Acceptable answers to a how-did-you-come-to-know question are determined by the nature and content of the proposition known, and this has a bearing on the distinction between empirical and *a priori* knowledge. To see that p is to know that p by empirical means. That makes one's knowledge empirical. To know that p by constructing a figure in pure intuition or, if there is such a thing, by rational intuition is to know that p by non-empirical means. That makes one's knowledge *a priori*. Since means of knowing are the key to the "what" question and some means of knowing yield empirical knowledge while others yield *a priori* knowledge one would expect an adequate answer to the what question to take account of the distinction between empirical and *a priori knowledge*.

But none of this changes the basic picture of knowledge for which I have been trying to make a case. Means of knowing, or of coming to know, remain at the centre of this picture and this is a reflection of the way in which attributions of knowledge are directly exposed to how-did-you-come-to-know questions and, in problematic cases, to how-possible questions. Yet it is armchair reflection rather than empirical science that exposes the links between “what”, “how”, and how-possible questions and it is armchair reflection rather than empirical science which reveals that all three questions can be answered by drawing on the notion of a means of knowing. Since another name for this kind of armchair reflection is “philosophical reflection” the methodological moral should be obvious: if we want to know what knowledge is and how it is possible there is no better way of proceeding than to do what I have been doing here: philosophy.[27]

REFERENCES

Austin, J. L. (1979). ‘Other Minds’, *Philosophical Papers*, Oxford University Press, Oxford.

Baldwin, T. (1990). *G. E. Moore*, Routledge, London.

Cassam, Q. (forthcoming). ‘Can the Concept of Knowledge be Analysed?’

Dray, W. (1957). *Laws and Explanation in History*, Oxford University Press, Oxford.

Dretske, F. (1969). *Seeing and Knowing*, RKP, London.

Gettier, E. (1963). ‘Is Justified True Belief Knowledge?’, *Analysis* 23/ 6.

Goldman, A. (1986). *Epistemology and Cognition*, Harvard University Press, Cambridge, Mass.

— (1992). ‘A Causal Theory of Knowing’, *Liaisons: Philosophy Meets the Cognitive and Social Sciences*, The MIT Press, Cambridge, Mass.

Kant, I. (1932). *Critique of Pure Reason*, trans. Norman Kemp Smith, Macmillan, London.

Kornblith, H. (1999). ‘In Defence of a Naturalized Epistemology’, *The Blackwell Guide to Epistemology*, (eds.) J. Greco and E. Sosa, Blackwell Publishers, Oxford.

— (2002). *Knowledge and its Place in Nature*, Oxford University Press, Oxford.

Longuenesse, B. (1998). *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the Critique of Pure Reason*, Princeton University Press, Princeton.

McDowell, J. (1998). ‘Singular Thought and the Extent of Inner Space’, *Meaning, Knowledge and Reality*, Harvard University Press, Cambridge, Mass.

[27] Thanks to Bill Brewer, John Campbell, David Charles, Ciara Fairley, Hanna Pickard, Declan Smithies, Timothy Williamson and audiences at Erfurt, Glasgow, Johns Hopkins, Oxford, Northwestern, Nottingham, Stirling, UCL, Utrecht, and Warwick for helpful comments on earlier versions of this material.

Moore, G. E. (1953). *Some Main Problems of Philosophy*, Allen & Unwin, London.

Nozick, R. (1981). *Philosophical Explanations*, Harvard University Press, Cambridge, Mass.

Quine, W. V. (1969). 'Epistemology Naturalized', *Ontological Relativity and Other Essays*, Columbia University Press, New York.

Strawson, P. F. (1997). 'Kant's New Foundations of Metaphysics', *Entity and Identity and Other Essays*, Oxford University Press, Oxford.

Stroud, B. (1984). *The Significance of Philosophical Scepticism*, Oxford University Press, Oxford.

— (2000). 'Scepticism and the Possibility of Knowledge', *Understanding Human Knowledge*, Oxford University Press, Oxford.

Williamson, T. (2000), *Knowledge and its Limits*, Oxford University Press, Oxford.

Grazer Philosophische Studien
74 (2007), 143–161.

TRANSCENDENTAL ARGUMENTS: A PLEA FOR MODESTY

Robert STERN
University of Sheffield

> ‘[If] any man were found of so strange a turn as not to believe his own eyes, to put no trust in his senses, nor have the least regard to their testimony, would any man think it worth while to reason gravely with such a person, and, by argument, to convince him of his error? Surely no wise man would’. (Reid 1863: 230)

Summary

A modest transcendental argument is one that sets out merely to establish how things need to appear to us or how we need to believe them to be, rather than how things are. Stroud’s claim to have established that all transcendental arguments must be modest in this way is criticised and rejected. However, a different case for why we should abandon ambitious transcendental arguments is presented: namely, that when it comes to establishing claims about how things are, there is no reason to prefer transcendental arguments to arguments that rely on the evidence of the senses, making the former redundant in a way that modest transcendental arguments, which have a different kind of sceptical target, are not.

Although it has never been an issue quite at the centre of recent epistemology, the promise and potential of transcendental arguments has received a good deal of discussion.[1] Up until now, the main focus of that discussion has been whether such arguments work, and in particular whether they can be used successfully to establish certain facts about the world in a sceptic-proof manner, where Barry Stroud’s influential paper from 1968 has persuaded many that they cannot.[2] I want to suggest here, however, that the argument of this paper is not as persuasive as is widely supposed.

[1] For a bibliography, see Stern 1999: 307–321.

[2] I will refer to the reprinted version in Stroud 2000: 9–25. Stroud’s was not the only critical voice from this period: another significant critic of transcendental arguments was Stephan Körner; see for example Körner 1967.

Nonetheless, I will claim, there is another way of criticising transcendental arguments to the same effect, which is to show that our hopes for such arguments should be modest and not ambitious. I will begin by sketching what I take transcendental arguments to be, and then consider Stroud's critique of them, before criticising this and offering an argument for modesty of my own. That argument hinges on whether there is any reason to prefer a transcendental argument as a response to scepticism over some other sort of response, for example, one that relies on the evidence of our senses? I will argue that in the ambitious way these arguments are commonly conceived, against the sceptic who is commonly taken to be their target, there is no reason to so prefer them; but if our conception of these arguments is made more modest, there is room to think they offer us something additional to other anti-sceptical manoeuvres, and so that it is here that their main value should be seen to lie.

I.

While the exact nature of transcendental arguments is far from unproblematic, they are generally taken to have the following features: They begin from some sort of self-evident starting point concerning our nature as subjects (for example, that we have experiences of a certain kind, or beliefs of a certain kind, or make utterances of a certain kind) which the sceptic can be expected to accept, and then proceed to show that this starting point has certain metaphysically necessary conditions, where in establishing that these conditions obtain, the sceptic is thereby refuted. So, in the face of the sceptical suggestion that we do not know that there is an external world, or other minds, or the past, a transcendental argument might be offered to provide deductive support for these claims from certain facts about our nature as subjects, based on the premise that the former are necessary conditions for the latter, where the form of the argument is: we have certain experiences etc; a necessary condition for us having these experiences etc is the truth of *S*; therefore *S*.

However, it is now widely held that this kind of argument is more problematic than it may at first appear. A highly influential source of this suspicion is Barry Stroud, who in his article 'Transcendental Arguments' suggested that for any claim concerning the necessary condition *S*, "the sceptic can always very plausibly insist that it is enough [that] we *believe* that *S* is true, or [that] it looks for all the world as if it is, but that *S* needn't actually be true" (Stroud 2000: 24).[3] So, in the case of the problem of the external world, for example, the concern is that no argument can be constructed to show that there must actually *be* an external world, but just that there must appear to us to be one, or that we must believe there to be one.

In subsequent work, Stroud has gone on to explain why the sceptic can "very plausibly" weaken the necessary condition for experience etc from *S* to "we must believe *S*" or "*S* must appear to be true". For, while he allows that we might reasonably be able to make modal claims about "how our thinking in certain ways necessarily requires that we also think in certain other ways", he thinks it is puzzling "how … truths about the world which appear to say or imply nothing about human thought or experience" (for example, that things exist outside us in space and time, or that there are other minds) "[can] be shown to be genuinely necessary conditions of such psychological facts as that we think and experience things in certain ways, from which the proofs begin". Stroud goes on: "It would seem that we must find, and cross, a bridge of necessity from the one to the other. That would be a truly remarkable feat, and some convincing explanation would surely be needed of how the whole thing is possible" (Stroud 2000: 158f).[4] Thus, Stroud is prepared to allow (and indeed exploits our capacity himself, in his own arguments against the sceptic)[5] "that we can come to see how our thinking in certain ways necessarily requires that we also think in certain other ways, and so perhaps in certain other ways as well, and we can appreciate how rich and complicated the relations between those ways of thinking must be" (Stroud 2000: 158f); but he believes that anything more than this, which asserts that "non-psychological facts" about the world outside us constitute necessary conditions for our thinking, is problematic.

[3] In this 'Transcendental Arguments' paper, the starting point for which *S* is meant to be a necessary condition is language, where if the transcendental claim could be established, *S* could be shown to be true from the fact that what the sceptic says makes sense. But, as many transcendental arguments have been proposed that are not just focussed on the conditions for language, I take it that Stroud's worry here cannot just apply to transcendental arguments of this sort, but also those that focus on conditions for experience, self-consciousness etc.

[4] Cf. also Stroud 2000: 212: "All this would be so on the assumption that transcendental arguments deduce the truth of certain conclusions about the world from our thinking or experiencing things in certain ways. That strong condition of success is what I continue to see as the stumbling-block for such ambitious transcendental arguments. Can we ever really reach such conclusions from such beginnings? … [The most troubling danger is] that of not being able to reach substantive, non-psychological truths from premises only about our thinking or experiencing things in certain ways".

[5] See, for example, Stroud 2000: 165 ff.

Faced with Stroud's challenge, it has appeared that there are three ways to go. First, one can opt for idealism, which sees no gap to bridge between how we think and how things are insofar as the former determines the latter, and then try to qualify this position in such a way as to render it somehow plausible. Second, one can opt for verificationism, which stipulates that some of what we believe about the world must be true. But both these options are seen as problematic in themselves, and as sufficiently anti-sceptical on their own to make any appeal to transcendental arguments redundant. A third possibility, however, is to opt for what can be seen as a more "modest" approach. On this view, it has been suggested that we accept that transcendental arguments should not attempt to cross Stroud's "bridge of necessity" at all; instead, we should allow that the only necessary conditions that we can establish concern how we must think or how things must appear to us, thus avoiding Stroud's call for an explanation of how we can get from the "psychological" to the "non-psychological" by remaining within the former, and eschewing claims about the latter. Stroud's own position viz-à-viz the sceptic involves this sort of modest approach, although there are other ways to take it.

However, before adopting any such "modest" strategy, the question should now be asked: how powerful is Stroud's position here? I think it is less compelling than is generally supposed.[6] For, according to Stroud, there is something inherently problematic in making a modal claim about how *the world* must be as a condition for our thought or experience, but there is not anything particularly problematic about making a claim about our thought or experience being a condition for some other aspect of our thought or experience. But why should it be somehow easier to make modal claims between ways of thinking or types of experience, than ways of thinking or types of experience and the world? Why are such "bridges" or modal connections easier to make "within thought" than between how we think and how the world must be to make that thought possible? Of course, one might take this symmetry between the two to be reason to be suspicious of modal claims of this sort at *any* level: but as we have seen,

[6] Whilst it is commonly taken as a starting point, Stroud's position has of course not gone totally uncriticised: see e.g. Glock 2003: 37–39. Glock cites an anticipation of Stroud's position from the lectures of C. D. Broad: "What Kant claims to prove by his transcendental arguments is that certain propositions, such as the law of causation and the persistence of substance, are *true* with the interpretation and within the range he gives them. But it is doubtful whether his arguments could prove more than that all human beings must *believe* them to be true, or must act *as if* they believed them to be true" (Broad 1978: 15). Broad gives no reason to substantiate his doubt here.

Stroud himself seems to think they are viable between "psychological facts".[7] If so, I believe, he needs to give us some account of why they are less problematic here than between our thought and the world; but as far as I can see he just takes it to be obvious, and so provides no such account.

Perhaps, however, it could be said that Stroud is right to take this to be obvious: for, if we are dealing with modal connections between us and the world, and the world is conceived in a realist manner, as independent of us, then of course we know less about how we depend on the world than on how we depend on other facts about us—this is just a feature of the mind-independence of the world, which makes it more opaque to us in this way, so that our modal claims concerning it are correspondingly more problematic, than they are concerning connections between our ways of thinking.

Now, of course, it is often taken to be the case that self-knowledge (knowledge of our "inner states") is less problematic than worldly knowledge, on these sorts of grounds, and Stroud's position may be taken to be trading on that intuition. But if so, I think it is mistaken. For, even if we grant that certain sorts of self-knowledge are unproblematic in this way, on grounds of their immediacy or self-evidence or infallibility, it seems unlikely that the modal claims involved in transcendental arguments would have these features, even if they involve merely "psychological facts" rather than claims about the world. If we make some sort of claim regarding the dependence of one aspect of our thought or experience on some other aspect of our thought or experience, what reason have we for thinking this is somehow self-evident or immediate in the way that (perhaps) "I am in pain" is self-evident and immediate—even though the former as well as the latter involve "facts about us" rather than the world outside us? It seems implausible that in the transcendental case, we have some sort of privileged "first person access" of the sort that might be used to establish a relevant epistemic difference in the non-transcendental case between "I am in pain" and "You are in pain", as what is involved in the transcendental case is not introspection but the use of modal intuition. Once again, therefore, there seems no reason to support Stroud's view concerning the asymmetry between ambitious and modest transcendental arguments, and so his claims for the greater viability of the latter over the former.

[7] Cf. Stroud 2000: 224–244, where Stroud defends the possibility of modal knowledge but still only considers modal claims involving "psychological facts", such as that "thinking of the world as contingently containing subjects of experience would require thinking of it as contingently containing objective particulars independent of experience as well" (236).

It appears, then, that we have little reason to accept Stroud's concerns that a special "bridge of necessity" is needed if we are to make transcendental claims that take us outside "psychological facts". If Stroud's position is to be rejected in this way, however, does that mean we should go back to conceiving of transcendental arguments in more ambitious terms? I do not think so, as I believe there is another reason that can be given for modesty, which is more compelling than the one offered by Stroud.

II.

This argument for modesty is really rather simple, and relates to the dialectic with the sceptic. The central thought is this: It is implausible to think that the sceptic would be satisfied with the use of any transcendental argument against him, given what has driven him to be a sceptic in the first place. We must either prevent him being driven to scepticism at some earlier point, in which case an appeal to transcendental arguments will be unnecessary; or we must accept that we cannot so prevent this, but by then a transcendental argument will come too late.

In order to see this, let me begin by characterising the sort of sceptic that an ambitious transcendental argument is supposed to convince, where I will focus on the problem of the external world. This sceptic is someone who has her reasons for doubting the truth of what most (or maybe all) of us believe we know: namely, that there is an external world of material objects outside us in space and time. This doubt is based on the thought that the kind of evidence we would present in favour of this knowledge is inadequate, where I take it that a large part of that evidence is perceptual. There are of course a number of arguments that the sceptic thinks she can give to show that this evidence is inadequate in this way, but most hinge on some sort of argument from error: we just cannot be sure that this evidence is sufficient to support our belief, given the compatibility of that evidence with various so-called "sceptical scenarios", such as Descartes' evil demon or the brain-in-a-vat hypothesis.

Now, it can be tempting to think that a transcendental argument is just what we need here, because it can seem that our difficulty is a fallibilist one: we must admit to the sceptic that our perceptual experience is fallible, so we cannot rule out the sceptical scenario on the basis of how things appear to us, as that appearance could be radically misleading. If the transcendental argument can be made to work, however, it has the advantage of being a *deductive* strategy, and thus error-proof—where it can be taken for granted that a sceptic who questions our reliance on the laws of logic has taken a step too far, even for a sceptic.

The difficulty, however, is that while a transcendental argument is indeed a deductive argument, it relies on certain premises which involve *modal claims*, about the world having to be a certain way in order for our experiences or thoughts to be possible. The question is, therefore, how can we satisfy the sceptic of our entitlement to make such modal claims?

Now, as much of the recent literature in this area suggests, our capacity to make such claims is somewhat mysterious—so mysterious, in fact, that even some who are not in any way sceptics would argue that we should eschew them.[8] However, my point here is not that general: let us assume that we do make such claims in a way that can be satisfactorily understood, so that we should have no doubts about their legitimacy on that score. Nonetheless, my worry is that if transcendental arguments rely on these modal claims, it is hard to see how they could then put us in a dialectically advantageous position with respect to the sceptic: for if the sceptic thinks she can question our perceptual evidence for the existence of an external world, can't she on very similar grounds question our modal intuitions for such claims of necessity—viz. that they are equally prone to error? For if (as seems plausible) we rely on criteria like conceivability or imaginability to test such modal claims, can't the sceptic plausibly say that our capacities here can go wrong, to the same degree as in the perceptual case—so how can the use of such claims make us better off?

One way to respond to a worry of this kind has been suggested recently by Thomas Grundmann and Catrin Misselhorn. They have argued that although a transcendental argument relies on our modal intuitions, this is not problematic, because the sceptic relies on such intuitions as well, in claiming that her sceptical scenario is metaphysically possible. They write: "If the sceptic claims that modal intuitions are unreliable, sceptical hypotheses could not be justified. For this reason, the sceptic must grant the reliability of modal intuitions as a method of justification" (Grundmann and Misselhorn 2003: 211). Thus, Grundmann and Misselhorn think they can get our modal intuitions to speak in favour of a claim like "Necessarily, perceptual beliefs about the external world are largely true", which can then be used as a premise of a transcendental argument to the

[8] The literature in this area is growing steadily. For a useful collection with a good selection of articles, see Gendler and Hawthorne 2002.

effect that we have perceptual knowledge of the external world, in a way that refutes the sceptic in an ambitious manner (cf. Grundmann and Misselhorn 2003: 207).

This is an interesting approach; I am not sure, however, that it properly does justice to the dialectic with the sceptic. It is certainly true that in appealing to sceptical scenarios, the sceptic tries to exploit the fact that he can make them seem metaphysically possible to us, where he uses our modal intuitions to do so. Thus, when our intuitions are going his way, so to speak, he is happy to exploit them. But suppose that Grundmann and Misselhorn are right, and that our intuitions can be made to go the other way, in support of a claim like "Necessarily, perceptual beliefs about the external world are largely true". What is to prevent the sceptic now changing his tune, and questioning our reliance on these intuitions? The sceptic, after all, is not someone with a settled position of his own to support: he will use whatever means are at his disposal to generate doubt—and if it turns out that using our modal intuitions is not an effective way to do so, because in the end they do not support his sceptical scenarios, why shouldn't he abandon them?

Of course, if he gives up appealing to his sceptical scenarios, the sceptic will still need *something* to base his doubt upon, and thus will now need some reason to question Grundmann and Misselhorn's transcendental argument and the modal intuitions it relies upon. But we have already seen what the ground of that doubt could be: namely, the claim that there is a possibility of error in the intuitions they use in support of their modal claim, such that the proof must remain open to question.

Now, Grundmann and Misselhorn might reasonably respond to this by saying: all *this* sceptical argument amounts to is an argument from fallibilism, and that "[i]t is generally accepted that fallibilism is not sufficient to generate scepticism" (Grundmann and Misselhorn 2003: 210), so the mere fact that our modal intuitions about this modal claim *might* be wrong is not a reason to doubt it—the sceptic has to have "proper modal intuitions speaking against it, and, as we have argued, the sceptic has not provided convincing modal evidence for his claim, so far" (Grundmann and Misselhorn: 218).

However, if we adopt this sort of strategy in defence of the use of transcendental arguments, what is to prevent us adopting it *from the beginning* of our debate with the sceptic, and apply it to the perceptual case: that is, why can't we dismiss merely fallibilistic arguments against our *perceptual* evidence for the existence of the external world, and ask the sceptic to come up with "proper evidence speaking against it"—for example, that there are brains-in-vats experiments going on, that scientists are available to conduct such experiments, and so on? In the absence of such evidence, why won't the "straight" perceptual evidence that we have for the existence of an external world do—making any appeal to a transcendental argument redundant in our response to the external world sceptic? To put this in the Reidian terms of my epigraph, by trying to use a transcendental argument against the sceptic, aren't we doing what no wise man should?

I think similar considerations would tell against another way of using transcendental arguments ambitiously. This would be to use them not in support of claims about the world in the face of arguments from error, but in support of claims about the reliability of our belief-forming methods, in the face of arguments from circularity—namely, we must presuppose such methods in order to establish their reliability.[9] Either we should block such scepticism from the beginning,[10] or the transcendental argument comes too late: for if we offer a transcendental argument to the effect that (for example) perceptual experience is reliable, as this reliability is a condition for something else, the question remains, how can we establish the reliability of the modal intuitions we employ in constructing the transcendental argument? Once again, therefore, it would seem that an appeal to transcendental arguments will not help in this situation.

If I am right, then, it turns out that there is no reason to think that the use of transcendental arguments gives us any advantage over the sceptic as he has been conceived so far: either he can be defeated some other way,[11] or if not, transcendental arguments gives us no additional advantage against him.

[9] For a useful discussion of the issues, see Alston 1993.

[10] For an attempt to do so, see Stern 2003: 229–232.

[11] Stroud, however, is not in a position to accept the argument against ambitious transcendental arguments sketched in this section, which is perhaps why he resorts to the argument against them outlined in section I. For, while he accepts that it is a central feature of the sceptic's position that he raises doubts based on possibilities for which no ground can be given, Stroud thinks we cannot reject scepticism simply on that score, as we could in the case of "ordinary" inquiries—where the sceptic's "extraordinariness" is on Stroud's view a corollary of the "extraordinariness" of the epistemological project itself. Ultimately, then, Stroud counsels that we should not try to answer the sceptic, but question the epistemological project that makes scepticism possible. See, for example, 'Taking Scepticism Seriously' and 'Understanding Human Knowledge in General', both reprinted in Stroud 2000.

III.

It might be argued, however, that weaknesses in the kind of non-transcendental, fallibilistic response I have given to the sceptic will lead to ambitious transcendental arguments being required after all. For, it could be said, these fallibilistic responses work by claiming that because the sceptic cannot give us any grounds for actually taking his sceptical scenario seriously and thinking that we are brains in vats etc., these scenarios pose no epistemic threat to our ordinary beliefs, as if they are mere logical possibilities, they do not provide us with sufficient reasons for doubting those beliefs. However, the sceptic could respond by arguing that even if he cannot give us any reasons for thinking that his sceptical scenarios are actually the case, they are still a threat, despite the fact that he cannot provide any evidence in their favour. I will here consider three sceptical strategies that might seem to undercut the fallibilist response to scepticism in this way, and whether or not these strategies require us to appeal to ambitious transcendental arguments if we are to deal with them.

A first sceptical strategy that seems to require no positive evidence in favour of the sceptical scenario I will call *the simple tracking argument*. On this strategy, it is claimed that the sceptical scenarios show that a belief like "I have two hands" fails to meet a fundamental tracking or sensitivity requirement on knowledge: viz. that if *p* was false, *A* would not believe that *p*. To show that I would fail to meet this requirement for knowing "I have two hands", the sceptic does not have to show that it *is* false or that I have good reason to think it is because I have good reason to think I am a brain in a vat: he just has to show that if I were a brain in a vat, it would be false, but I would continue to believe it, thereby showing (he argues) that my belief "I have two hands" violates the tracking requirement for knowledge.

In the face of this sceptical position, it may seem tempting and indeed obligatory to return to some sort of ambitious transcendental argument. For, in response to the tracking problem, it could be claimed on the basis of such an argument (of the sort suggested by Putnam, for example)[12], that the sceptic is wrong to suggest that in the sceptical scenario my belief

[12] See Putnam 1981: Ch. 1. Putnam himself takes as his target the very coherence of the sceptical hypothesis; but I am here suggesting that some of his claims about reference could also be used to resolve the tracking problem, on the grounds that if I were a brain in a vat my belief "I have hands" would have a different "vatted" meaning, and so not fail to track how things are, because how things are would shape the content of my belief.

"I have two hands" would fail to track the truth, because I could not be a brain in a vat while still believing that I have hands, as my belief would have a different reference, so that the tracking condition on this belief (and others like it) can be guaranteed not to fail in this manner.

In fact, however, I think the fallibilist has a perfectly adequate response to the sceptic here, without recourse to any ambitious transcendental argument of this kind. Of course, if the fallibilist is not to avail himself of an argument like Putnam's, he must allow that if he were a brain in a vat, he would continue to believe that he has hands, and thus that the tracking requirement would fail in this respect. But the question here is whether it is a plausible condition on knowledge that if *p* were false, *A* would not believe that *p* in *all* situations in which *p* might be false? What makes it implausible that this condition holds is precisely what makes infallibilism implausible: namely, that we can know things, even when our grounds for knowing them or our methods for knowing them are prone to lead to error in some circumstances. What matters, of course, is in *what* circumstances we are prone to error: if we are error-prone in circumstances that would require a lot of manipulation[13] in order for us to make the error (such as would be needed to make the brain in a vat scenario work) then the fact that we could not track the truth in these circumstances arguably does not count against our knowing the truth in more normal ones. Rather than telling against fallibilism, therefore, all the tracking objection reveals is one of its consequences, namely that knowledge does not require a logical entailment between not-*p* and *A* not believing *p*:[14] it all depends on how and why that relation breaks down, where it can be claimed that in the case of the sceptical scenario, that breakdown would not be enough to show that our capacity to track "I have two hands" is inadequate for knowledge.

It appears, then, that the simple tracking argument can be defeated without any need to appeal to an ambitious transcendental argument. However, this strategy focused on our belief "I have two hands", and

[13] There are different ways this idea could be worked out. One could be in terms of possible worlds, namely, that is only in worlds some distance from the actual one that we will be fooled. Another way might be in terms of normal functioning, namely, our cognitive mechanisms would require serious distortion for us to be misled. And of course, depending how "normal functioning" is spelt out, these approaches may end up converging.

[14] Cf. Nozick 1981: 199, who observes that the tracking condition we have been considering "does not say that in all possible situations in which not-*p* holds, S doesn't believe *p*. To say there is no possible situation in which not-*p* yet S believes *p*, would be to say that not-*p* entails not-(S believes *p*), or logically implies it".

directly claimed (implausibly as it turned out) that that belief fails to meet the tracking requirement, because it would not do so in a sceptical scenario. However, the sceptic might now offer a more complex tracking argument, which I will call the *tracking plus closure argument*. Rather than focusing on the belief "I have two hands", this argument focuses on the belief "I am not a brain in a vat" and claims that *this* fails the tracking argument; it then follows from the closure principle that "I have two hands" is not known.

The first step in this strategy, then, is to claim that I do not know I am not a brain in a vat because this belief fails to track: if I were a brain in a vat, I would still believe I am not. But why can't we respond to this sceptical point as before: namely, why can't the belief "I am not a brain in a vat" be said to only fail to track that it is false in exceptional circumstances, namely when we are envatted, just as the belief "I have two hands" only fails to track the truth in such circumstances? If this failure doesn't prevent us knowing we have hands in the latter case, how does it prevent us knowing that we are not brains in vats in the former? However, I think there is something different in the brain in a vat case which blocks this sort of response: for, while there are circumstances in which I would track the falsity of "I have two hands" (for example, if they were chopped off in an accident, or as a result of disastrously incompetent surgery) but just fail to do so in extreme circumstances like being envatted, the sceptic can claim that there are *no* circumstances in which I would track the falsity of "I am not a brain in a vat"—because if it were false, and so I was a brain in a vat, then *ex hypothesi* I would never pick this up and so never change my belief accordingly.[15] Thus, while an externalist might say that in both cases my belief only fails to track in a remote possible world, I would be prepared to grant the sceptic that these beliefs are not on a par, and that a successful tracking argument can be made against my belief "I am not a brain in a vat".

The second step of the sceptical argument is then to go from this admission to the conclusion that I do not know I have two hands, via the closure principle: if *A* knows *p*, and *A* knows that *p* entails *q*, then *A*

[15] So, it will not do here to say that I might come to pick up the falsity of "I am not a brain in a vat" in some circumstances, for example if the evil scientist were not very competent and sent me some information that tipped me off. I am taking it that it is part of the sceptic's *conception* of what it is to be a brain in a vat that the scientists concerned never make such errors. If such malevolent perfection seems implausible to attribute to human scientists, substitute evil demon scientists instead.

knows *q*. Using modus tollens on this principle, it can be argued that as we don't know we are not brains in vats, then we don't know we have two hands either (as it seems right to say that if we have hands, this entails we are not brains in vats). It would seem, then, that even without giving us any reason to think we actually are brains in vats, the sceptic can use his scenario to undermine our ordinary beliefs, such as the belief that we have hands.

Now, of course, there are a variety of responses to this problem in the literature which do not employ any sort of transcendental argument strategy, such as approaches that deny the closure principle, or that reject the tracking requirement on knowledge, replacing it with a weaker requirement that our belief "I am not a brain in a vat" can be said to meet.[16] The fallibilist can therefore explore ways out of this difficulty without being obliged to adopt a transcendental argument. But still, it might be suggested, if a transcendental argument can be used against the sceptic here without requiring any such manoeuvres, that might seem to show that transcendental arguments have a significant role to play against the sceptic in enabling us to answer him without modifying what appear to many to be plausible epistemic principles.

I would claim, however, that the promise of transcendental arguments in this respect is once again illusory. It may seem that the way in which such an argument could be used is in relation to the tracking issue. For, it could be said, arguments such as Putnam's show that the tracking requirement for "I am not a brain in a vat" can be met, by showing that this is something we could not falsely believe: for, if I were a brain in a vat, thinking "I am not a brain in vat" would not refer to brains in vats but something else (perhaps vat images).[17] The difficulty is, however, that although this sort of transcendental argument meets the tracking requirement, it can nonetheless be plausibly claimed that there are further conditions it does not meet. For, it doesn't show what also seems to be needed, which is that the grounds on which we form the belief "I am not a brain in a vat" do not prevent us from believing it when it is false; the transcendental argument just shows that semantic externalist conditions on reference make it impossible to believe "I am not a brain in a vat" falsely, much as physical conditions on belief make it impossible to believe "I am alive" or "There is oxygen in the room" falsely, without in itself showing that

[16] For the former approach, see e.g. Nozick 1981: 197–247, and Dretske 1970. For the latter approach, see e.g. Sosa 1999.

[17] Putnam 1981: 14f.

the grounds we have for that belief are what make us sensitive to its truth and falsity—and this seems to be what is required for knowledge. If I am right, we will therefore be obliged to find other ways of responding to the sceptical challenge such as those mentioned above, that do not employ transcendental arguments of this sort.

The same sort of difficulty applies to another way of using a transcendental argument in relation to the tracking issue, which is even stronger than the semantic externalist one we have just considered. This would be to try to meet the tracking problem by using a transcendental argument to show that "I am not a brain in a vat" is akin to a necessarily true mathematical proposition[18]: for, it could be said that the sceptical scenario is metaphysically impossible, because we could not be brains in vats, as being envatted would prevent us from being believers at all. If this is right, then it would seem that the sceptical scenario is necessarily false, as it could never be actualised; so just as "if 2 + 2 = 4 were false" is a necessarily false supposition, doesn't this show that the same is true of supposing the falsity of "I am not a brain in vat"?

The difficulty here, however, is that all this argument would show is not that the tracking requirement has been *met* by my belief "I am not a brain in a vat", but rather that the antecedent of the tracking conditional is necessarily false, as is also the case for necessary truths like 2 + 2 = 4; but then, as Nozick has argued, it seems that it is best to say not that the tracking requirement has been satisfied and so that our belief constitutes knowledge in this respect, but that it is not a requirement at all (Nozick 1981: 186f). Now of course, even if "I am not a brain in a vat" cannot turn out to be falsely believed by me, this does not in itself show that I know it, any more than the fact that '2 + 2 = 4' cannot be falsely believed by me shows I know it either: I still need some adequate grounds for believing each of them, where it is precisely those grounds that the sceptic questions, in the transcendental argument case as much as in the ordinary perceptual one. We therefore still need to be told why the sceptic should take the transcendentalist's reasons for believing he is not a brain in a vat more seriously than the non-transcendentalist's.

Finally, we can consider a third sceptical strategy that again treats the sceptical scenario as a mere possibility, which might be called *the priority argument*. Here, the sceptic claims that there must be a certain *order* to

[18] Of course, it can't be exactly the same, as I do not exist in all possible worlds, whereas numbers (arguably) do.

our knowledge, such that to know familiar things like "I have two hands" I must *already know* things like "I am not a brain in a vat". But then, the sceptic argues, on what grounds could I know that I am not a brain in a vat, if all the empirical premises from which I might infer this can only be accepted *after* the hypothesis has been refuted? Unless we are prepared to follow G. E. Moore, and argue from these empirical premises directly against the sceptical hypothesis, it seems like we are required to argue against it while being deprived of any basis on which to do so, where again the sceptic here does not need to provide any positive evidence in its favour in order to make his case.

Once again, in the face of this difficulty, it may seem that only something like an ambitious transcendental argument can give us what is needed: for, it gives us an argument against the sceptical scenario which is a priori, and which can therefore be used to *first* show that we are not brains in a vat, in a way that seemed to be required before we could lay claim to any of our ordinary empirical knowledge. We might therefore grant the sceptical suggestion that we must know that we are not brains in vats before we can know that we have hands etc., but employ a transcendental argument to show that this requirement can be met.

However, the problem with this way of using transcendental arguments is that the sceptic's priority argument is less than compelling, despite being plausible on the surface. For, the crucial move is to say that we cannot know an empirical proposition like "I have two hands" unless we already know that the sceptical hypothesis is false and can rule it out. But why should we accept this move? Consider the following propositions:

a: my copy of *War and Peace* is in my study

and

b: my copy of *War and Peace* has not been stolen.

Do I have to have evidence for *b* before I can come to know *a*? The answer would seem to depend on the circumstances. If I already know (or have reason to believe) that a lot of stealing of Russian classics has been going on, then even if I clearly remember putting *War and Peace* in my study this morning, have a generally good memory etc., that may not be sufficient grounds for knowing *a*, unless I am in a position to rule out *b*. But in different circumstances, where as far as I know stealing of this sort never or very rarely happens, why must I establish *b* before my evidence for *a* can be accepted? It seems to me this is not required; and of course, the fallibilist argues that we find ourselves in this latter situation when it comes to the sceptical scenario, where we have no grounds for believing this scenario to hold. If this is the right approach to take (as I believe it is) it undercuts the need to argue against the sceptical scenario in an a priori manner using a transcendental argument as a first step: we can just claim (as before) that the lack of evidence for the sceptical scenario is sufficient to justify us in accepting our ordinary empirical beliefs for the usual reasons, such as perceptual evidence, memory, testimony and so on.

IV.

It appears, then, for reasons rather different from those presented by Stroud, but more to do with the dialectical situation involved in our debate with the sceptic, that ambitious transcendental arguments have little work to do.

Does that mean that *all* forms of transcendental argument have little work to do however—modest ones included? I would suggest not.[19] For, while it turns out that transcendental arguments have little to add in the battle against the external world sceptic who thinks our perceptual evidence is insufficient because error-prone, this is not the *only* way to be an external world sceptic. For example, one can argue that the perceptual content of our experience does not tell us anything about an external world as we believe it to be, but that we get such beliefs by inferring from a more impoverished perceptual content (as on Hume's view that "It's commonly allow'd by philosophers, that all bodies, which discover themselves to the eye, appear as if painted on a plain surface, and that their different degrees of remoteness from ourselves are discover'd more by reason that the senses" [Hume 1978: 56]), where the issue then is how such inferences can be justified. In this situation, I have argued elsewhere[20], a transcendental argument that concerns merely the perceptual content of our experience (how things *appear to us*) can be useful, in making our beliefs concerning the external world direct and perceptual rather than indirect and inferential.

[19] For further discussion see Stern 2000.

[20] See Stern 2000: Ch. 4.

Now, if we adopt a target of this kind for our transcendental argument, I think we can avoid the dialectical difficulty we faced previously. Previously, it seemed dialectically inappropriate to use modal claims against the sceptic who argues from error, when the grounds for making such claims seem at least as vulnerable to error as the empirical grounds (such as perceptual experience) that these claims are supposed to replace; but there is less dialectical incongruity in making transcendental claims against the kind of sceptic I have just sketched. For in the latter case, we really *are* in the sort of situation viz-à-viz the sceptic that Grundmann and Misselhorn hoped we were in previously, as the sceptic is interested in doing more than offer a mere argument from error and fallibility to question our belief; she is trying to show not just that our evidence is fallible, but that the belief is not justified in our own terms, given the problematic nature of the inference from how she takes things to appear to us to how we think they are, where the "veil of appearances" seems to mean that "reason" could not have enough to go on in making any such inference. Similarly, therefore, the sceptic must do more than just offer an argument from fallibility against a transcendental claim that how things appear to us must be rich enough to support our belief in the external world on direct perceptual grounds (for example, contra Hume, that the world is immediately presented to us in three dimensions, so that this is not "discover'd more by reason that the senses"); instead, he must give us grounds for thinking that this claim is not properly supported by our modal intuitions, where then he must show that experience on his impoverished model would be sufficient to be a condition for us to be the kind of conscious creatures we are, contra our transcendental argument against him. The sceptic is thus not in a position to make a *general* argument against our reliance on our modal intuitions to support our transcendental claim about how our perceptual experience must be, so if our intuitions can be made to speak in favour of this claim, he cannot shrug them off as he did previously, but must show why in this case our intuitions are in fact mistaken, or can be made to go the other way. The dialectic of this situation, then, gives the transcendental argument some genuine work to do, with some prospect of success—but only a transcendental argument of a modest kind.

If I am right, however, that a transcendental argument can only be used successfully in this modest way, against this form of sceptic, what about the sceptic of the more radical kind, who argues from various sceptical scenarios, and against whom it seemed that an ambitious transcendental argument might be needed? If we have abandoned the latter form of argument, does this sceptic therefore win the day? And if so, if we use a transcendental argument to defeat the more modest sceptic, can't she remain a sceptic by turning more radical?

In abandoning the use of an ambitious transcendental argument against the radical sceptic, however, all that was claimed was that there is no reason to think these arguments should be any more successful than approaches that do not use such arguments, some of which have been mentioned above.[21] Thankfully, therefore, if I am right to say that we must live without ambitious transcendental arguments at this level, and be content to settle for modesty, I think it is also right to say that we can do so without conceding defeat to the sceptic, of either the more modest or the more radical kind.[22]

REFERENCES

Alston, W. P. (1993). *The Reliability of Sense Perception*, Cornell University Press, Ithaca and London.

Broad, C. D. (1978). *Kant: An Introduction*, Cambridge University Press, Cambridge.

Dretske, F. (1970). 'Epistemic Operators', *Journal of Philosophy* 67, 1007–1023.

Gendler, T. S. and Hawthorne, J. (eds.) (2002). *Conceivability and Possibility*, Oxford University Press, Oxford.

Glock, H.-J. (2003). 'Strawson and Analytic Kantianism', *Strawson and Kant*, (ed.) H.-J. Glock, Oxford University Press, Oxford, 15–42.

Grundmann, T. and Misselhorn, C. (2003). 'Transcendental Arguments and Realism', *Strawson and Kant*, (ed.) H.-J. Glock, Oxford University Press, Oxford, 205–218.

Hume, D. (1978). *A Treatise of Human Nature*, L. A. Selby-Bigge (ed.), 2nd edn., rev. P. H. Nidditch, Oxford University Press, Oxford.

[21] Of these, the one I would favour would be of a quasi-pragmatist kind, hinted at above: The sceptic cannot give us any positive reason to think that the sceptical scenarios actually hold, and as such cannot use them to generate any real doubt, of the sort that might provide a genuine threat to our beliefs about the world.

[22] I would like to thank Alex Burri and Christian Beyer for inviting me to the "Philosophical Knowledge" conference, for which this paper was originally written; and I am grateful to participants at the conference for comments on that occasion, and particularly to Ernest Sosa for correspondence on it afterwards. I am also grateful for comments from Paul Faulkner, Bob Hale and Christopher Hookway, and to participants at the departmental seminar of the University of York, where a previous version was also delivered.

Körner, S. (1967). 'The Impossibility of Transcendental Deductions', *Monist* 61, 317–331.

Nozick, R. (1981). *Philosophical Explanations*, Oxford University Press, Oxford.

Putnam, H. (1981). *Reason, Truth and History*, Cambridge University Press, Cambridge.

Reid, T. (1863). 'Essays on the Intellectual Powers of Man', *The Works of Thomas Reid*, (ed.) W. Hamilton, 2 vols, 6th edn., Maclachlan and Stewart, Edinburgh, 215–508.

Sosa, E. (1999). 'How to Defeat Opposition to Moore', *Philosophical Perspectives* 13, 141–153.

Stern, R. (ed.) (1999). *Transcendental Arguments: Problems and Prospects*, Oxford University Press, Oxford.

— (2000). *Transcendental Arguments and Scepticism*, Oxford University Press, Oxford.

— (2003). 'On Strawson's Naturalistic Turn', *Strawson and Kant*, (ed.) H.-J. Glock, Oxford University Press, Oxford, 219–234.

Stroud, B. (2000). *Understanding Human Knowledge*, Oxford University Press, Oxford.

Grazer Philosophische Studien
74 (2007), 163–175.

A PRIORI EXISTENCE

Alex BURRI
Universität Erfurt

Summary

This paper deals with the question whether existence claims may be supported in an *a priori* manner. I examine a particular case in point, namely the argument for the existence of so-called logical atoms to be found in Wittgenstein's *Tractatus*. Although I find it wanting, I argue that more general reflections on the notion of existence allow us to straightforwardly answer our initial question in the affirmative.

1. *Introduction*

Are there things whose existence can be justified *a priori*? As always in philosophy, it depends on whom you ask. Think, for example, about the argument for the existence of a substantial ego—whether this presumed substance should really be regarded as immaterial is a separate issue—that Descartes put forward in the second Meditation and, especially, in the first chapter of his *Principia Philosophiae*.[1] This argument can be considered *a priori* because it does not rely on any particular experience or any specific experiential content. The mere fact that we do have experiences, of whatever content, suffices for it to go through. However, the very same kind of entity Descartes took to be indubitable has been rejected by Hume and Wittgenstein for the reason that we are unable to encounter substantial egos in experience.[2,3] Here, as elsewhere, one philosopher's *modus ponens*

[1] In contrast to the *Meditations*, the *Principia Philosophiae* contain an explanation of why the ego is a substance as opposed to, say, a mere bundle or sequence of thoughts. See Descartes 1644: 8.

[2] "For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe any thing but the perception" (Hume 1739: 252; his emphases).

[3] "There is no such thing as the subject that thinks or entertains ideas. / If I wrote a book

is another philosopher's *modus tollens*.

In this paper I shall examine another potential candidate, namely Wittgenstein's aprioristic argument to the effect that logical atoms must exist. I hope that doing so will help us to acquire a better understanding of the notion of existence and its kin.

2. *Varieties of existence*

Existence comes in a *de re* and a *de dicto* mode. If we use notation based on the lambda-calculus, taking the singular term "$\lambda x.Fx$" to denote the property of *F*-ness, we can render the *de re* interpretation of "existence" as

$$\lambda x.\exists y(y = x).$$

According to this reading, existence turns out to be a philosophically unexciting property since the question "'What is there?' [...] can be answered [...] in a word—'Everything'—and everyone will accept this answer as true" (Quine 1948: 1). Indeed, the claim "Everything exists" is true *a priori* since "$\forall x \exists y(y = x)$" is a theorem of first-order predicate logic with identity. If existence is to be of any concern, we have to switch from ontology in the narrow sense to what Quine calls ideology, especially to the question what *kinds* of things there are.

This leads us to the *de dicto*—or *de conceptu*, to be more precise—reading which goes back to Frege's contention that existence, like number, is a second-order property of concepts—namely the property of being instantiated. It cannot straightforwardly be expressed in the familiar idiom; the closest first-order representation we can get is

$$\lambda x.\exists y(y \in x),$$

which stands for the set-theoretic property of being non-empty. Existence so understood does cut philosophical ice since there are not only entities (concepts or first-order properties) that have it but also entities that lack it.

called *The World as I found it*, I should have to include a report on my body, and should have to say which parts were subordinate to my will, and which were not, etc., this being a method of isolating the subject, or rather of showing that in an important sense there is no subject; for it alone could *not* be mentioned in that book" (Wittgenstein 1922: § 5.631; his emphasis).

Now, what can be known *a priori* is often equated with what we can extract from our concepts, by conceptual analysis as it were. As Frege was eager to point out, however, what can be extracted from a concept are its features (*Merkmale*), i.e. the traits that enter into its definition, but not its (second-order) properties. I think the reason is the following: since a concept may or may not be instantiated, the property of being instantiated must surely be inessential to the identity conditions of the concept. Being instantiated, then, is not part and parcel of conceptual content[4] although Frege does not categorically exclude the possibility that this property might somehow be implied by certain conceptual features.[5]

This is not to say that *de dicto* existence eludes every aprioristic treatment. For example, "$\exists x \neg \exists y (y \in x)$", the null set axiom of Zermelo-Fraenkel set theory, is trivially known to be true *a priori* since it is introduced by a stipulation.[6] Nonetheless, the property of being instantiated remains a foreign, maybe even irrelevant, addition to the features controlling both the inferential role and the non-inferential application of a concept. Hence, it would come as a surprise if *de dicto* existence manages to sustain any aprioristic investigation—whereas its *de re* complement, albeit deployable independently of experience, completely lacks substance.

For metaphysical purposes, however, the *de dicto* variant is not as useful as one might wish—even if one has already abandoned hope of carrying out an *a priori* analysis. This becomes clear, I think, when we consider predicates like "is abstract", "is fictional", "is mythical" or "is unreal" which are themselves used to mark off ontological differences. Since such predicates also indiscriminately express concepts that may or may not be instantiated, *de dicto* existence, taken by itself, cannot mirror the ontological distinctions we are intuitively inclined to make. Or, to put it somewhat differently, if the property of being fictional is exemplified by Sherlock Holmes, then *de dicto* existence *per se* is powerless to mark off the metaphysical dissimilarity between Sherlock Holmes and, say, Bill Clinton.

There are simply too many properties for *de dicto* existence to get the metaphysical work done all on its own. Quite to the contrary, the metaphysical burden seems to rest almost exclusively on the shoulders of the concepts (or kinds of concepts) themselves. In order to carry some of the

[4] Conceptual content for Frege is tied to inferential role. See Frege 1879, 2 f. [§ 3].

[5] See Frege 1884: 86 f. [§ 53] and Frege 1990: 21.

[6] Of course, a set of interconnected stipulations may turn out to be inconsistent, thus undermining the initial truth claim—witness Cantor's set theory and Russell's antinomy. Proofs of inconsistency, however, are themselves *a priori*.

weight *de dicto* existence therefore needs refinement or supplementing. Without restrictions on the scope of "*F*", that is to say, existence *qua F* will fail to amount to existence *in the relevant sense*—whatever that may mean. Hence, distinguishing between concepts whose exemplification guarantees "relevant" existence and concepts whose exemplification remains ontologically inconsequential becomes vital. However, such a distinction cannot be drawn on a purely logico-semantic basis. Rather, it requires a substantial metaphysical theory which, in turn, threatens to transform any philosophical investigation—at least if it is intended to be carried out *a priori*—into a *petitio principii*.

The difference between (unrestricted) *de dicto* existence and what I offhandedly called "existence in the relevant sense" is reminiscent of a distinction made by Russell long before he harshly criticized Meinong for taking advantage of a similar idea—namely the distinction between "being" and "existence":

> *Being* is that which belongs to every conceivable term, to every possible object of thought—in short to everything that can possibly occur in any proposition, true or false […]. Being belongs to whatever can be counted. […] "*A* is not" must always be either false or meaningless. […] Numbers, the Homeric gods, relations, chimeras and four-dimensional spaces all have being, for if they were not entities of a kind, we could make no propositions about them. Thus being is a general attribute of everything […]. / *Existence*, on the contrary, is prerogative of some only amongst beings. To exist is to have a specific relation to existence (Russell 1903: 449).

As the remark "being is a general attribute of everything" suggests, the open-minded notion of being, which resembles my unrestricted *de dicto* existence, is on the brink of collapsing into sheer *de re* existence. So we are left with Russell's enigmatic characterization of "existence" that is every bit as noncommittal as the phrase "existence in the relevant sense". But we can shed some light on it by making what looks like a detour: from the epistemic notion of apriority I now turn to the metaphysical notions of necessity and contingency.

3. *A puzzle about existence*

Existence is a paradigmatic case of contingency. That sulfides contain sulfur or that dolphins are mammals may be necessary but that there are any such things as sulfides or dolphins certainly isn't. Whether some particular thing or kind of thing exists always depends on other things—and dependence is a distinguishing feature of contingency. Although unanswerable, Leibniz's famous question why there is something rather than nothing does not lack sense. For, that there might be nothing seems conceivable. Consequently, the existence of every single thing must be contingent, so-called necessary beings such as God or pure sets notwithstanding.

The deeper reason for this contingency lies in the peculiarity that existence, in whatever mode, fails to be a genuine property. (If it were a genuine property, it would be possible for things to differ from each other merely in the presence or absence of that property; by Leibniz's law such things would be non-identical, hence two in number; but in order to be two, both would have to exist, which contradicts the initial assumption.) But by failing to be a genuine property, existence cannot be included in the essence of a thing, as would be required if the latter's being were necessary.[7,8] Therefore, existence is always contingent—or so it seems.

Problems lurk around the corner, however. What does it mean to say that an object exists contingently? The existence of an object, say *a*, is said to be contingent if *a* might not have existed, i.e. if it is possible that *a* does not exist. According to the familiar idiom, this in turn amounts to the claim there is at least one possible world in which *a* (or a counterpart thereof) does not exist. Hence, *a*'s contingency *presupposes* the existence of a certain possible world, i.e. possibility. In consequence, the existence of such a possibility or possible world cannot, on pain of an infinite regress, be contingent.[9] The realm of possibilities, the totality of possible worlds or the so-called logical space must have a different modal status than ordinary objects—otherwise it could not provide for a framework for embedding modal discourse.

Next, consider the possible worlds themselves. We saw that they con-

[7] This is, very roughly, Kant's refutation of the ontological argument for God's existence as presented in his first *Critique* (see B 620–30).

[8] It is sometimes claimed that the standard definition of an essential property—an object has a property essentially if it would cease to exist without it—makes existence an essential property of everything (see e.g. Forbes 1997: 516; Plantinga 1995: 139; Kripke 1971: 87, n. 11). But if the definition is restricted to real properties, this conclusion doesn't follow.

[9] To put it in David Lewis's words: "We think of the totality of all possible worlds as if it were one grand world, and that starts us thinking that there are other ways the grand world might have been. […] But this is thoroughly misguided. If I am right, the many worlds already provide for contingency, and there is no sense in providing for it all over again. […] There is but one totality of worlds; it is not a world; it could not have been different" (Lewis 1986: 80).

tain, or do not contain, ordinary objects such as *a*, Bill Clinton, and Mt. Kilimanjaro (or counterparts of them). It is, therefore, advisable to just identify a possible world with the mereological sum of all the objects which are parts of it.[10] But now there is trouble. For a possible world has every such part essentially. Just as "it is essential to the identity of a set that it have the members that it does" (Fine 1981: 179), a possible world cannot contain other individuals than it does—or else it would be a different world. However, if possible worlds both exist non-contingently and have their parts essentially, then the latter must exist necessarily. And this contradicts the contingency of existence from which we started out.

The simplest solution to the problem is to plead ambiguity: "regular" modal claims about mundane objects would have to be distinguished from "extraordinary" modal claims about matters concerning logical space itself. According to that proposal, a regular modal sentence such as "It is possible that unicorns exist" expresses *world-restricted* content and can be interpreted in the standard way, whereas an extraordinary modal sentence such as "It is possible that a plurality of worlds exists" involves *unrestricted* quantification and must, therefore, be understood differently.[11] Thus necessary existence in the regular sense could be strictly constrained to existence in all (accessible) possible worlds—a requirement that a commonplace object like Mt. Kilimanjaro or Pegasus still does not meet, quite independently of the question whether possible worlds and their parts exist non-contingently in some other sense of the modal term.

Another solution consists in rejecting the identification of possible worlds with mereological sums of concrete, run of the mill individuals. Instead, one could regard possibilities in general and possible worlds in particular as abstract entities, for example as (nested) bundles of properties, themselves to be considered as universals. According to such a view, the actual world is the only possible world in which those property-bundles are instantiated—in which, that is to say, the abstract, purely qualitative "way the world is" is exemplified in concrete, substantial matter. This dissolves the puzzle because now the existence of a possible world no longer implies the existence of concrete particulars. Thus the latter do not inherit the presumed non-contingency of the former.

We shall come back to both suggestions later on. For the moment, however, it is worth emphasizing that the very prospect of having to deal with

[10] See Lewis 1986: 69.

[11] This is, in effect, the solution that John Divers has proposed. See Divers 2002: 47–9, where he also proposes some principles governing extraordinary modality.

necessary existence—not merely in one case but across the board—reopens the apriority issue. And there is indeed an example of presumed *a priori* existence most relevant to our discussion.

4. *A Tractarian argument*

In the *Tractatus*, Wittgenstein puts forward a rather enigmatic argument to the effect that certain entities—i.e. simple, non-composite, and unchanging "objects" that "form the substance of the world"—must exist: "If the world had no substance, then whether a proposition had sense would depend on whether another proposition was true" (Wittgenstein 1922: § 2.0211). This observation, as condensed as it may be, has the structure of a transcendental argument purporting to demonstrate that something (namely, the existence of logical atoms) is the unavoidable precondition of something else (namely, there being meaningful descriptions of the world). It consists of two claims: first, the meaningfulness of a sentence cannot depend on the truth of another sentence; second, the independence of meaning from truth presupposes the existence of simple objects.

This argument has received differing interpretations,[12] but it seems clear that it is intended to be *a priori* in the following sense: the proclaimed existence of logical atoms is not supposed to depend on there being those complex, ordinary objects that we actually encounter in experience. In that respect, Wittgenstein's argument diverges significantly from the one Leibniz is giving at the very beginning of his *Monadology*: "There must be simple substances, since there are composite ones" (Leibniz 1714: § 2; my translation). In Wittgenstein's case, the reason given for the atomistic conclusion is semantic, not empirical. It stems from considerations concerning the proper functioning of language *qua* representational medium.

In a nutshell, the first part of the argument may be reconstructed as follows: We notice, with Quine, "that truth in general depends on both language and extralinguistic fact. The statement 'Brutus killed Caesar' would be false if the world had been different in certain ways, but it would also be false if the word 'killed' happened rather to have the sense of 'begat'" (Quine 1951: 36). In consequence, the fact that a statement means such and such, i.e. has such and such a sense, cannot depend on the truth of

[12] Cf. e.g. Glouberman 1980: 26f. and Skyrms 1993: 221f.

any (other) statement—at least not systematically, or else a vicious circle or an infinite regress would ensue.

To put it otherwise, the sense of a sentence either is or represents a truth condition: "Instead of, 'This proposition has such and such a sense', we can simply say, 'This proposition represents such and such a situation'" (Wittgenstein 1922: § 4.031). And since it is, according to Wittgenstein, "the agreement or disagreement of its sense with reality" that "constitutes [the] truth or falsity" of the sentence (*ibid.*: § 2.222), the sense does indeed specify the condition under which the sentence is (or would be) true. Whether this condition is met or not, is then determined by extralinguistic fact. But the specification of the condition must already be in place *before* we can establish the statement's truth value. That is why meaning takes epistemological priority over truth—it is *a priori* by definition while truth is not.

Such precedence guarantees the independence of meaning from truth, as required by Wittgenstein's first claim. But his second claim still remains to be argued for. So why should this independence presuppose the existence of simple objects? Let us consider a statement like "Unicorns are timid": it must have a sense that, at least in principle, enables us to decide on the statement's truth or falsity. Vagueness notwithstanding, this involves having a sense specific enough to divide all objects we encounter in experience into complementary groups such as unicorns and non-unicorns or timid and non-timid things, respectively.[13] If we were incapable of determining with some reliability whether some animal in front of us is a unicorn or not, then the truth value of the statement would completely escape us.

But since the sense of "Unicorns are timid" does not depend on the truth of "There are unicorns",[14] the sense of "unicorn" cannot derive from what Russell calls "knowledge by acquaintance". Our ability to distinguish between unicorns and non-unicorns cannot, that is to say, stem from a history of causal interactions with certain animals in which we developed "the capacity noninferentially to respond appropriately and differentially" (Brandom 2000: 21) to such things as unicorns. Rather the sense of "unicorn" must amount to a more or less complex description of these putative entities. However, such a description can only serve its purpose—i.e.

[13] Such a differentiation does not have to be infallible, however. Occasionally taking fool's gold for gold or an elm for a beech does not impair one's *linguistic* competence as long as one is prepared to accept an expert's correction.

[14] This corresponds to the example Wittgenstein gives in his 'Notes Dictated to G. E. Moore in Norway'.

to help us telling *F*s and *non-F*s apart—if its components, ultimately, do refer to things that we know by acquaintance.

A description that aspires to be applicable in experience must, that is to say, exploit recurring elements of experience. For if its components were themselves descriptions consisting of still further descriptions, without end, one could neither apply nor understand it. So eventually the description has to consist of *semantically* simple components, called "names" by Wittgenstein. A name in this *quasi*-Kripkean sense has no descriptive content; it merely "stands for one thing" (Wittgenstein 1922: § 4.0311; see also § 3.221). The last step of the argument must then consist in identifying the referents of those semantically simple components with our logical atoms, i.e. with the *ontologically* simple entities we were looking for.

But what could legitimize such a move? Why should the recurring elements of experience in which our descriptions are ultimately anchored themselves be simple? One reason is this: If the recurring elements were of high or even infinite complexity, we could not readily recognize them—and speedy detection certainly is a precondition of any effective application of language. And if their complexity were low, then nothing could (and indeed would) prevent us from dismantling them into their truly basic parts or aspects. So we end up with non-composite entities whose existence has been bolstered by a more or less *a priori* argument.

5. *Pasting the pieces together*

If a sentence is meaningful, then all its constituent parts have sense, too. And since the meaningfulness of "Unicorns are timid" does not depend on there being unicorns, the predicate "unicorn" must have a sense that transcends all the common objects we encounter in experience. But how can that be—given that the sense in question has the traits of a *description*? What is it a description of? The answer which suggests itself is: the sense of the word "unicorn" portrays *possible* objects, or "possibilia" for short.

Since such an object must be describable independently of whether we have encountered it in experience or not, its description is, so to speak, *ante rem*. According to the Wittgensteinian picture just sketched, this can only be managed if the description appeals, ultimately, to semantically simple expressions whose denotation is known by acquaintance. In consequence, possibilia must themselves be complex, composite entities made up of logical atoms, i.e. of the things designated by those simple expressions. Besides their semantic function logical atoms do therefore also have a metaphysical one: they are the elementary building blocks out of which possibilia are (or can be) put together.

Indeed, Wittgensteinian atoms have a "logical form" that enables them to combine with each other in certain ways, thus giving rise to more complex things: single states of affairs, whole situations, and the everyday, run of the mill objects we encounter in experience. Since such complexes are assembled entities, their *possibility* is entirely determined by the logical—read: combinatory—form of the simple objects they are (or could be) composed of: "If all objects are given, then at the same time all *possible* states of affairs are also given;" logical atoms, therefore, "contain the possibility of all situations" (Wittgenstein 1922: §§ 2.0124 and 2.014, respectively; his emphasis).

From this it follows that every possible or "imagined world, however different it may be from the real one, must have something—a form—in common with it" (*ibid.*: § 2.022). This common "form" has two aspects: first, a set of unchangeable logical atoms and, second, a catalog of logical or combinatorial laws enabling their assembly. From a Tractarian point of view, possible worlds differ from each other neither in the fundamental entities they are composed of nor in the laws governing their building but rather in the way the logical atoms are combined with each other. Possibilities in general and possible worlds in particular are *recombinations* of these elementary building blocks.

Taken together, the logical atoms and the logico-combinatorial laws *generate* logical space. And since they are the very source of all possibilities, they themselves cannot be contingent. Moreover, this metaphysical point has an epistemological analogue: because the meaning of an empirical sentence must be grasped before its truth value can be determined (with the aid of experience), and because sentence meaning coincides with the description of a possible state of affairs, we have *a priori* access to these possibilities.

What we do not know *a priori* is, of course, which possibilities are actual, i.e. how the logical atoms are in fact combined. Meaning or sense, we have seen, has to be descriptive in order to permit verification; it must, that it to say, allow a comparison with extralinguistic reality by specifying the truth condition of the sentence in question. Truth conditions, in turn, are nothing but "*possible* situations" or "*possible* states of affairs" (Wittgenstein 1922: §§ 2.0122 and 2.0124, respectively; his emphases), namely those that would have to obtain, i.e. to be actual, for the respective sentences to be true. Whether the sense of a sentence corresponds to the facts, i.e. to an obtaining situation, depends on how the (actual) world is. Describing a possible state of affairs makes a statement meaningful, while depicting an actual state of affairs makes it (empirically) true.

The Wittgensteinian solution to our puzzle about existence (see § 3) is analogous to the actualist response mentioned above: it avoids equating possibilia with *concrete* individuals (and possible worlds with mereological sums thereof). According to the Tractarian model, the only concrete particulars are the logical atoms and their factual configuration, while their possible but non-actual recombinations do only exist *in abstracto*. This solution has the huge disadvantage, however, of interfering in physics' most intimate business. For in order to save the non-contingency of the possibilities, Wittgenstein must claim to have an *a priori* justification for the existence of particulars that are *both concrete and actual*—and that is too much of a good thing. For example, it precludes physics from having fields as fundamental entities.

It is therefore preferable to stick with David Lewis's possibilism, i.e. with the view that possible worlds and their concrete parts do all exist *sui generis*. So possibilities do not have to be generated (by actual means); they are already there anyway. And because Lewisian worlds do neither overlap nor causally interact with each other, possibilism does not impose any restriction on the constituent parts—ultimate or otherwise—of the actual world. Thus it avoids patronizing physics from the outset. But how, then, is the puzzle about existence to be solved?

We plead ambiguity. In the basic, *de re* sense literally everything exists, including Pegasus—and non-contingently so. For, in order to exist contingently logical space, i.e. the sum of all possibilities, would have to contain an alternative to itself as a proper part, which would lead, Löwenheim-Skolem notwithstanding, to incoherence; and what is true of logical space as a whole is transmitted to its parts (the possible worlds) and subparts (the possibilia) as well.

In the technical sense, necessary existence means having a counterpart in every possible world—a privilege no concrete object seems to possess. Contingent or necessary existence in this technical sense depends, so to speak, on how widespread one's family is. Hence, it is about *size* rather than existence (in the basic, *de re* sense).

Finally, what about the notion of "existence in the relevant sense" that we got as an ill-defined substitute for Frege's unfortunate *de dicto* existence? What is "the further privilege of existence" (Russell 1903: 71) that befalls only some of the things there *are* (in Russell's sense of "being")? It is, I claim, nothing else than actuality: to exist in the relevant sense is to be actual, i.e. to be a part of the actual world. But which possible world is the actual one? Well, *this* one, of course. In other words, the answer to that question is indexical and irreducibly *de se*.[15] To be actual is to belong to the world that contains *my consciousness*. Hence, actuality is something intrinsically relational rather than something intrinsic like existence (again, in the basic sense).

Are there things whose existence can be justified *a priori*? Surprisingly enough, everything's existence seems to be justified in that way. However, we should not confuse existence with two properties more dear to us, namely being widespread across possible worlds (whose ideal limit is omnipresence or necessary existence in the technical sense) and being actual.

REFERENCES

Brandom, R. B. (2000). *Articulating Reasons*. Cambridge, Mass.: Harvard University Press.

Descartes, R. (1644). *Principia Philosophiae*, in *Œuvres de Descartes*, vol. VIII, (eds.) C. Adam and P. Tannery. Paris: Vrin 1996.

Divers, J. (2002). *Possible Worlds*, London: Routledge.

Fine, K. (1981). 'First-Order Modal Theories I—Sets', *Noûs* 15, 177–205.

Forbes, G. (1997). 'Essentialism', in *A Companion to the Philosophy of Language*, (eds.) B. Hale and C. Wright. Oxford: Blackwell, 515–33.

Frege, G. (1879). *Begriffsschrift*. Hildesheim: Olms 1993.

— (1884). *Grundlagen der Arithmetik*. Stuttgart: Reclam 1987.

— (1990). 'Dialog mit Pünjer über Existenz', in his *Schriften zur Logik und Sprachphilosophie. Aus dem Nachlass*. Hamburg: Meiner, 1–22.

Glouberman, M. (1980). '*Tractatus*: Pluralism or Monism?', *Mind* 89, 17–36.

Hume, D. (1739). *A Treatise of Human Nature*, (ed.) L. A. Selby-Bigge. Oxford: Clarendon Press 1978.

Kripke, S. (1971). 'Identity and Necessity', reprinted in *Metaphysics. An Anthology*, (eds.) J. Kim and E. Sosa. Malden: Blackwell 1999, 72–89.

[15] "Likewise for the fact of which world is ours. It is an egocentric fact, on a par with the fact of which person is me—in fact, the latter subsumes the former" (Lewis 1986: 130; see also 92f.).

Leibniz, G. W. (1714). 'Monadologie', in his *Philosophischen Schriften*, vol. 6, (ed.) C. I. Gerhard. Hildesheim: Olms 1978, 607–23.

Lewis, D. (1986). *On the Plurality of Worlds*. Malden: Blackwell.

Plantinga, A. (1995). 'Essence and Essentialism', in *A Companion to Metaphysics*, (eds.) J. Kim and E. Sosa. Oxford: Blackwell 138–40.

Quine, W. V. (1948). 'On What There Is', in his *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press 1961, 1–19.

— (1951). 'Two Dogmas of Empiricism', in his *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press 1961, 20–46.

Russell, B. (1903). *The Principles of Mathematics*. London: George Allen & Unwin 1937.

Skyrms, B. (1993). 'Logical Atoms and Combinatorial Possibility', *The Journal of Philosophy* 90, 219–32.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*, trans. by D. F. Pears and B. F. McGuinness. London: Routledge 2001.

Grazer Philosophische Studien
74 (2007), 177–197.

SELF-REFERENTIAL ARGUMENTS IN PHILOSOPHY

Elke BRENDEL
Universität Mainz

Summary

The paper discusses the strengths and weaknesses of arguments of proper self-reference, arguments of self-application and arguments of iterative application. A formalization of the underlying logical structure of these arguments helps to identify the implicit premises on which these arguments rest. If the premises are plausible, the conclusions reached by these arguments must be taken seriously. In particular, all the types of argument discussed, when sound, show that certain theories that purport to be universally applicable are not tenable. The argumentative power of such arguments then depends on how devastating it is for the theories in question to give up their claim of universal applicability.

1. *Introduction*

Self-referential arguments are highly regarded in philosophical disputes. They are often considered to be particularly profound and strongly persuasive. Self-referential arguments are typically destructive in nature. They are often employed in order to show that a philosophical theory leads to a contradiction—or at least to seriously counterintuitive results—with respect to certain self-referential propositions the theory allows one to construct. In logic and formal semantics, there are paradoxes yielded by self-referential arguments—like Russell's paradox and the famous Liar paradox—that have had an enormous impact on our views about fundamental concepts in set theory and on our view about the concept of truth, respectively. But also in other philosophical disciplines, self-referential arguments are often used to undermine some of the most basic assumptions of a theory. In spite of their ubiquity in philosophical reasoning, self-referential arguments are rarely the topic of metaphilosophical examination. Most textbooks on informal logic or critical thinking do not even address the strengths and weaknesses of these kinds of arguments.[1]

In the following, I will argue that there is not just one single type of self-referential argument, but rather that we have to distinguish between at least two main types of self-referential argument: arguments of *proper self-reference* and arguments of *self-application*. I will also argue that arguments of self-application need to be divided into three subversions: arguments of *propositional self-application*, arguments of *predicative self-application*, and arguments of *individual self-application.* I will furthermore analyze another type of argument which I refer to as arguments of *iterative application*. Although these latter arguments are not strictly speaking self-referential arguments, they share some important logical features with arguments of self-application.

After a formalization of the underlying logical structures of all these arguments, I will examine their argumentative power. In particular, I will identify the implicit premises on which these arguments rest and discuss the intuitive plausibility of these premises. If the premises are plausible, the conclusions reached by such an argument must be taken seriously. In particular, all the types of arguments discussed, when sound, show that certain theories that purport to be universally applicable are not tenable. The argumentative power of such arguments then depends on how devastating it is for the theories in question to give up their claim of universal applicability.

2. *Arguments of proper self-reference*

Arguments of proper self-reference employ a sentence A which is said to be equivalent to a sentence that attributes or denies a certain property to the very same sentence A. It is then shown that this self-referential structure generates an inconsistency or reveals counterintuitive results of a given theory.[2] The structure of arguments of proper self-reference can be put

[1] Holm Tetens' recent book *Philosophisches Argumentieren*—see Tetens 2005—is an exception here. He examines only very briefly certain kinds of self-referential arguments and dismisses them as unpersuasive. According to Tetens, despite their "mass production" in philosophy, convincing self-referential arguments are extremely rare. In contrast to Tetens, I will make a more positive appraisal of the argumentative strength of self-referential arguments in this paper.

[2] The term "theory" is used throughout this paper in a rather lax way. It does not only apply to full-fledged philosophical systems, but also to less elaborated accounts or to single hypotheses.

more generally and formally as follows:

> *Formal structure of arguments of proper self-reference*:
> Let $B(y)$ be a formula of a given theory T in which just the variable y is free.
> Then the sentence $A \leftrightarrow B(\ulcorner A \urcorner)$ generates a contradiction or reveals seriously counterintuitive results of T. Therefore, T has to be rejected in its present form.
> (Where A is a sentence of T and $B(\ulcorner A \urcorner)$ is the sentence resulting from substituting y with $\ulcorner A \urcorner$ (the quotation name of A) in $B(y)$.)

In the famous Liar Paradox, for example, a sentence L "saying of itself" that it is not true, i.e, $L \leftrightarrow \neg \mathit{True}(\text{“}L\text{”})$, yields a contradiction in a theory of truth in which the Tarski convention of truth holds.[3] Since the Tarski convention of truth is usually regarded as a fundamental adequacy condition for truth, the Liar Paradox is widely considered to pose a serious threat to our basic intuitions about the concept of truth and has resulted in controversial debates about the correct theory of truth.

Similar arguments of proper self-reference are employed with regard to the concept of knowledge. In the so-called *Knower Paradox* which goes back to a paper by Kaplan and Montague (Kaplan/Montague 1960), a self-referential sentence together with very plausible premises governing the use of the concept of knowledge seem to provide a fatal contradiction. The first premise is the widely accepted assumption that knowledge implies truth:

(1) $\mathrm{K}_{\mathrm{S}}(\ulcorner A \urcorner) \rightarrow A$

for all sentences A, epistemic subjects S and the knowledge predicate K.

The second premise asserts that every epistemic subject S—or at least every epistemic subject S possessing a minimal conception of knowledge—knows that knowledge implies truth:

(2) $\mathrm{K}_{\mathrm{S}}(\ulcorner \mathrm{K}_{\mathrm{S}}(\ulcorner A \urcorner) \rightarrow A \urcorner)$, i.e. $\mathrm{K}_{\mathrm{S}}(\ulcorner \text{ (1) } \urcorner)$.

[3] According to Tarski's truth convention, all equivalences of the form "X is true if and only if p" should follow from an adequate and formally correct theory of truth (where p is a sentence of this theory and X is the name (the quotation name or a structural description) of this sentence—for details of Tarski's account of truth, see Tarski 1935.

Now consider the following self-referential construction of a sentence G "saying of itself" that it cannot be known by an arbitrary epistemic subject S:

(*) $G \leftrightarrow \neg\, K_S(\text{“}G\text{”})$.

From $\neg G$ and (*), we get $K_S(\text{“}G\text{”})$, from which together with (1) we get G. We have thus derived $\neg G \rightarrow G$—and therefore G. Since we have demonstrated a proof of G from premise (1), we *know* that G follows from (1), and since we know premise (1) according to premise (2), it seems to be very reasonable to assume that we also *know* that G, i.e. $K_S(\text{“}G\text{”})$. But from $K_S(\text{“}G\text{”})$ with (*) we get $\neg G$. We have thus derived the contradiction $G \wedge \neg G$.

Since we have derived this contradiction with intuitively plausible assumptions about knowledge, the Knower Paradox seems to have detected a serious flaw in our conception of knowledge. It seems that premise (1) and premise (2) are beyond reproach. Truth is generally considered as a necessary condition for knowledge, and an epistemic subject who is capable of understanding the concept of knowledge knows this conceptual fact about knowledge. In deriving $K_S(\text{“}G\text{”})$ which led us to the second conjunct of the contradiction, $\neg G$, we made use of an instance of the so-called *principle of epistemic closure*. According to this principle, if a subject S knows *p* and knows that *p* implies *q*, S also knows that *q*. There is an ongoing debate about the plausibility of this principle (Barke 2002; Hales 1995). I am not going to pursue this important question any further. However, it should be noticed, that even if this principle is not universally valid, the *instance* of this principle we used in the above reasoning seems to be perfectly okay. So, do we really have to give up plausible premises or a plausible principle of reasoning in order to solve the Knower Paradox?

Before answering this question, I would like to mention another argument of proper self-reference that attempts to show the impossibility of an omniscient being.[4] This argument also makes use of the sentence (*) that gives rise to the Knower Paradox: Suppose S* is an omniscient being. Then, for every sentence A, S* should know A if and only if A is true—and *vice versa*—, i.e.

(**) S* is an omniscient being iff: $A \leftrightarrow K_{S*}(\ulcorner A \urcorner)$ for all sentences A.

[4] See, e.g., Brendel 2001. Also see Tetens 2004: 85f.

Let's assume S* is omniscient. So, if *G* were true, S* would know that *G*. From $K_{S*}(\ulcorner G \urcorner)$ together with (*), it follows that $\neg G$; but if *G* were false, S* wouldn't know that *G*—which, according to (*), is exactly the meaning of *G,* i.e. *G* would be true. Since we have derived a contradiction, S* cannot be omniscient.

It should be noticed that a theist need not be impressed by this argument in the first place. Even if the assumption of omniscience results in a contradiction with self-referential sentences, S* can still know vastly more than any finite being like ourselves. S* can still know all empirical facts, and all semantic facts that do not lead to a contradiction. But what does the above argument of proper self-reference then really show?

In order to evaluate the strengths and weaknesses of the arguments of proper self-reference presented above, we must carefully examine the implicit assumptions that are necessary for the employment of these kinds of arguments. First and most important is the very simple but nevertheless widely overlooked point that the paradox-generating sentences, like (*), must—formally speaking—be *theorems of the given theory T*. If the sentence that results in a contradiction does not follow from the theory that is attacked by this argument, we can simply reject the sentence as false. Why on earth should we in this case believe in a sentence that was made up for the only reason to be inconsistent with the theory we believe in? Only if this sentence follows from the theory we should take the inconsistency seriously.

When do paradox-generating sentences, like (*), follow from a given theory? If we are dealing with formalized theories, this question has a precise answer. Kurt Gödel has shown (Gödel 1931) that a formal system *T* which is sufficiently strong to include the arithmetic of natural numbers "can speak" about its own expressions, sentences and proofs via a certain effectively calculable function—the "Gödel numbering" function—that assigns natural numbers to expressions, sentences and proofs in *T*. In particular, *T* can formally represent various of its own metatheoretic syntactic notions—like "provability". Within *T* the following *theorem* can be proved:[5]

Diagonal Lemma

For any formula $B(y)$ of *T* in which just *y* is free, there is a sentence *A* such that:

[5] For a proof of the diagonal lemma, see Boolos/Jeffrey 1989: 173f.

$T \vdash A \leftrightarrow B(\ulcorner A \urcorner)$.

So, if $\neg\, \mathrm{K}_{\mathrm{S}}(y)$ were a formula of T, there would be a sentence G such that (*) is a theorem of T and as a consequence the Knower Paradox would be indeed a serious threat to our conception of knowledge. But is $\neg \mathrm{K}_{\mathrm{S}}(y)$ a formula of T? First of all, there is a very simple point that should be taken into account: In order to get the self-referential sentence (*) via diagonalization, "knowledge" needs to be a *predicate*. Indeed, in (*) it is assumed that K_{S} functions as a predicate so that "“G”" in "$\neg\, \mathrm{K}_{\mathrm{S}}$(“$G$”)" is an individual constant. But in most accounts of knowledge and in particular in most systems of epistemic logic, "knowledge" is considered to be a sentence *operator*. If "knowledge" were a sentence operator, we would get the following equivalence instead of (*):

(*)′ $G \leftrightarrow \neg \mathrm{K}_{\mathrm{S}}(G)$,

(1) would then be read as follows:

(1)′ $\mathrm{K}_{\mathrm{S}}(A) \to A$ for all sentences A and epistemic subjects S,

and (2) would amount to:

(2)′ $\mathrm{K}_{\mathrm{S}}(\mathrm{K}_{\mathrm{S}}(A) \to A)$ for all sentences A and epistemic subjects S.

Since (*)′ is no longer a theorem of T, we can simply reject (*)′ as false—like any other assumption that leads to a contradiction, and no paradox gets off the ground.

Even if there are good reasons to treat "knowledge" as a predicate, the derived paradoxical result still rests on some disputable assumptions. In order to apply the diagonal lemma, "knowledge" need not only be a predicate of T, but also a syntactic predicate of T that can be formalized within T. If we treat "knowledge", like "truth", as a semantic predicate that cannot fully be defined within T, but only in a richer metalanguage of T whose knowledge-predicate can also only be defined in a meta-metalanguage, etc., the paradox disappears. In this case we have a hierarchy of different knowledge-predicates: On the lowest level, we have a knowledge-predicate that applies only to empirical sentences A involving no knowledge-predicate; on the next level, we have a knowledge-predicate applying to sentences of the form $K_{\mathrm{S}}(\ulcorner A \urcorner)$; on the next level, we have a knowledge-predicate applying to sentences $K_S(\ulcorner K_S(\ulcorner A \urcorner)\urcorner)$; and so on. A sentence (*) in this approach is not even a syntactically well-formed sentence, and so, again, no paradox can get off the ground.

If there are convincing reasons to consider "knowledge" as a *syntactic predicate* of T, there is still a final possible move available that would prevent the paradox: In treating K_S as a syntactic predicate of T, "$K_S(\ulcorner A \urcorner)$" can be interpreted as "A is known to be provable (in T by S)". The Knower Paradox would then amount to a proof of the fact that *if T is consistent neither G nor ¬G is known to be provable in T*—in other words, we would get a result analogous to Gödel's first incompleteness theorem, the latter of which is hardly inconsistent.[6]

The upshot of the above analysis of arguments of proper self-reference is this: Arguments of proper self-reference rest on various formal conditions. In particular, it has been shown that arguments of proper self-reference involving the concept of knowledge should be viewed as a serious objection to an account of knowledge, only if it can be made plausible that the paradox-generating sentence follows from this account of knowledge, which in turn requires that "knowledge" is a predicate satisfying all the formal conditions for diagonalization. Whether all these conditions are fulfilled with respect to the notion of knowledge is doubtful.[7]

3. *Arguments of self-application*

I now turn to another kind of self-referential argument, namely, *arguments of self-application.* I distinguish between arguments of *propositional* self-application, arguments of *predicative* self-application, and arguments of *individual* self-application.

3.1. *Arguments of propositional self-application*

In arguments of propositional self-application, a statement A of a theory T falls into the domain of application of A itself. So, statement A consti-

[6] For a detailed attempt at solving the Knower Paradox using this strategy, see Brendel 2004.

[7] The Knower Paradox is usually met with indifference by most epistemologists. My above diagnosis of the strengths and weaknesses of arguments of proper self-reference could be a vindication of such an attitude.

tutes an example of what A itself expresses. But in applying A to itself, an inconsistency or a counterintuitive result follows within T. To be more precise, the general structure of arguments of propositional self-application can be described as follows:

> *Formal structure of arguments of propositional self-application*:
> A is a statement (for example, a definition, an axiom, a law, a thesis) of a theory T and has the following form of a universal sentence:
> (A) For all x which are of kind K: $B(x)$.
> A is itself of kind K. Therefore, $B(\ulcorner A \urcorner)$. $B(\ulcorner A \urcorner)$ generates a contradiction or reveals seriously counterintuitive results of T. Therefore, T has to be rejected in its present form.
> ($B(x)$ is a formula in which x occurs free, and $B(\ulcorner A \urcorner)$ results from $B(x)$ by substituting $\ulcorner A \urcorner$ for x in $B(x)$.)

Arguments of propositional self-application are clearly the most popular self-referential arguments in philosophy. They arouse considerable attention in philosophical discourse, and philosophers usually refer to these kinds of arguments when they talk about self-referential arguments (e.g., Tetens 2004: 94).[8]

There are numerous examples of arguments of propositional self-application in philosophy. One famous example of an argument of propositional self-application is an argument that attempts to show that Logical Positivism is self-referentially incoherent. One main thesis of Logical Positivism is its *verification principle of meaning* (VPM). Roughly, this principle can be put as follows:

> (VPM) For all propositions p: p is cognitively meaningful iff p is either analytically true or false, or can be confirmed or disconfirmed by observation.

(VPM) itself is a proposition, but (VPM) is clearly neither an analytical proposition nor can (VPM) be confirmed or disconfirmed by empirical means. As a consequence, (VPM) must be considered as cognitively meaningless, i.e. as a pseudo-proposition which lacks a truth-value, according to the standards of meaningfulness that (VPM) itself lays down. Therefore,

[8] For Steven Yates, providing arguments of propositional self-application is a characteristic method of philosophy and can produce, if carried out properly, "genuine philosophical knowledge" (see Yates 1991: 134f.).

Logical Positivism seems to be seriously flawed, since according to its own standards of cognitive meaningfulness the main ideas of Logical Positivism cannot be described by cognitively meaningful propositions.

Another example of an argument of propositional self-application is Alvin Plantinga's argument designed to show that (classical) foundationalism is self-referentially incoherent (Plantinga 1983: 61ff. and Plantinga 2000: 93 ff.). According to classical foundationalism (CF), the standards for justified belief are the following:

> (CF) For all propositions *p* and (rational) persons S: S is justified in accepting a belief *p* if and only if *either* (i) *p* is properly basic for S (i.e., *p* is self-evident, incorrigible, or Lockeanly evident to the senses for S), *or* (ii) S believes *p* on the evidential basis of propositions that are properly basic and that evidentially support *p* deductively, inductively, or abductively (Plantinga 2000: 93f.).[9]

(CF) is a proposition, and if a classical foundationalist wishes to be rational, she must, of course, be rationally justified in accepting (CF). (CF) must therefore itself satisfy the classical foundationalist's standards of rational belief. But Plantinga now observes that (CF) itself is not properly basic for any epistemic subject S. He further contends that no classical foundationalist has ever produced an argument showing that there are properly basic propositions supporting (CF). Therefore, no person S is rationally justified in believing (CF). So, the standards for justified belief that are laid down by classical foundationalists themselves are not met by their own main claim about foundationalist justification, i.e. the main claim of classical foundationalism—namely (CF)—is accepted by classical foundationalists, although it is not rational for them to believe it according to their own standards of rational belief. Plantinga, of course, uses this argument of propositional self-application in order to reject (CF) and, thereby, open the door to the possibility of there being properly basic beliefs (like the belief in god) that are neither self-evident nor evident to the senses nor incorrigible for S.

Another prominent example of an argument of propositional self-

[9] A proposition *p* is, according to Plantinga, "Lockeanly evident to the senses for S" if and only if *S* immediately knows *p* by a sensation that is caused by "external objects of some kind or other; not that those objects have the properties of trees, horses, or the other sorts of objects we think there are" (Plantinga 2000: 93, FN 48).

application is the refutation of relativism (with respect to truth).[10] The main claim of most versions of relativism amounts to something like the following thesis (R):

> (R) For all propositions *p*: *p* is true if and only if *p* is true relative either (i) to a conceptual framework, or (ii) to a certain paradigm, or (iii) to whether most people in a respected scientific community take *p* to be true.

Since (R) is itself a proposition, (R) is according to the relativist's own standards, only relatively true, which in turn means that there could be a conceptual framework in which (R) is false, i.e. in which absolutism is true.

How convincing are arguments of propositional self-application? The argumentative persuasiveness of these kinds of arguments diminishes if a defender of a theory *T* can provide reasons to the effect that the main claim *A* of *T* can be exempted from the domain of the universal quantifier in *A*. It could be argued, for example, that the domain of the universal quantifier in (VPM), (CF) or (R) is restricted to propositions that are not metatheoretic in character, i.e. that are not propositions about propositions, so that these claims wouldn't be instances of the universal quantifications they themselves express. But there is no plausible rationale for this restriction. After all, there are numerous metatheoretic propositions that are clear cases of cognitively meaningful propositions which do not lead to problems for Logical Positivism. For example, the metatheoretic proposition (1) "For all propositions *p*: *p* = *p*", seems to be a clear case of an analytically true proposition, and the metatheoretic proposition (2) "There are propositions *p* and subjects S such that S does not believe that *p*" is a proposition that can be empirically confirmed. Thus, both of these metatheoretic propositions are cognitively meaningful according to (VPM). Moreover, examples (1) and (2) can be regarded as propositions supported by properly basic propositions which are therefore propositions that one can be justified in accepting, according to (CF). But neither (1) nor (2) lead to inconsistent results for classical foundationalism. For a classical foundationalist, (1) is self-evident;[11] (2) seems to be deductively supported by the proposition

[10] For a more detailed examination of this argument see, for example, Yates 1991: 135f. and Tetens 2004: 86f.

[11] Or, if universally quantified sentences are not considered to be self-evident, it can be argued that (1) is inductively supported by the self-evident propositions a = a, b = b etc.

that S′ does not believe p′ (for a specific subject S′ and a specific proposition p′) which a classical foundationalist can be justified in believing on the basis of abductively reasoning based on S's behavior. Finally, a full-blooded relativist could also regard (1) and (2) as propositions that are only relatively true. Therefore, reducing the domain of the universal quantifier so as to exclude *all* metatheoretic propositions would be too restrictive since it would rule out clear-cut cases of metatheoretic sentences that do not yield incoherent results on (VPM), (CF), or (R).

Another strategy to avoid the devastating consequences of arguments of propositional self-application is to bite the bullet and admit that (VPM) is not cognitively meaningful, that the classical foundationalist is not rationally justified in believing (CF), and that (R) is not relatively true according to the respective theories in question. But this, so the argument goes, does not undermine the theories, since (VPM), (CF) or (R) are only necessary for setting up the theory. They are means to formulate the theory, but as such they are not part of the theory. They are, as Wittgenstein describes it in his famous ladder-metaphor at the end of the *Tractatus*,[12] like a ladder leading us to the right theory, but after reaching it we have to throw them away since they cannot be regarded as propositions coherent with the theory.

I find it very hard to make sense of this Wittgensteinian strategy. If, as far as (VPM) is concerned, a central thesis of Logical Positivism is cognitively meaningless according to the very standards of Logical Positivism, how can it then lead to a theory that claims to be a serious semantic theory? A cognitively meaningless sentence does not have, according to Logical Positivism, any truth-value whatsoever and as such cannot express any proposition. Thus it seems that it cannot be used in any kind of reasoning in developing a semantic theory.

The above arguments of propositional self-application seem to me devastating, since they show that the main claims of the respective theories do not satisfy the standards laid down by their own theories. To exempt (VPM), (CF) or (R) from its own domain of application in order to avoid the inconsistent result, strikes me as not only *ad hoc*, but also as a serious weakening of the original positions.

[12] See Wittgenstein 1921: 6.54 "My propositions are elucidatory in this way: he who understands me finally recognizes them as senseless, when he has climbed out through them, on them, over them. (He must so to speak throw away the ladder, after he has climbed up on it.)"

3.2 *Arguments of predicative self-application*

In arguments of *predicative* self-application not the whole statement A, but a *predicate-term that occurs in A* falls into the domain of application of A. This form of self-application then leads to an inconsistency within the theory in question. The logical structure of arguments of predicative self-application can thus be stated as follows:

> *Formal structure of arguments of predicative self-application*:
> A is a statement (for example, a definition, an axiom, a law, a thesis) of a theory T and has the following form of a universal sentence:
> (A) For all x which are of kind K: $B(P(x))$.
> P is itself of kind K. Therefore, $B(P(\ulcorner P \urcorner))$. $B(P(\ulcorner P \urcorner))$ generates a contradiction or reveals seriously counterintuitive results of T. Therefore, T has to be rejected in its present form.
> ($B(P(x))$ is a formula in which x occurs free, P is a predicate constant, and $B(P(\ulcorner P \urcorner))$ results from $B(P(x))$ by substituting $\ulcorner P \urcorner$ for x in $B(P(x))$.)

Grelling's paradox of heterological predicates is a famous example of an argument of predicative self-application (Grelling/Nelson 1908). A predicate is called *heterological* if and only if it doesn't apply to itself:

> (HET) For all (one-place) predicates x: x is heterological if and only if x does not possess the property expressed by x.

So, so for example, "monosyllabic" is heterological, since "monosyllabic" is a polysyllabic term and is therefore not true of itself, whereas "polysyllabic" is not heterological (but rather autological) because "polysyllabic" is polysyllabic.

Since "heterological" is itself a (one-place) predicate, it falls into the domain of the universal quantifier in (HET). We can therefore substitute "heterological" for "x" in the above definition:

> "Heterological" is heterological if and only if "heterological" does not possess the property expressed by "heterological".

That "heterological" does not possess the property expressed by "heterological" means, of course, that "heterological" is not heterological, i.e.:

> “Heterological” is heterological if and only if “heterological” is not heterological,

Although the concept of a heterological predicate defined by (HET) seems to be perfectly intelligible, the above argument of predicative self-application nevertheless shows that (HET) leads to a contradiction within the framework of a classical semantic theory in which, in particular, an interpretation function assigns the property of being *X* to the predicate term “*X*” as its semantic value.

In contrast to the above discussed arguments of propositional self-application, there seem to be more plausible reasons to the effect that the domain of the universal quantifier in (HET) should be restricted to predicates of the *object-language*. Consequently, so (HET) as originally formulated cannot be correct. Thus, if one wishes to provide a coherent definition of “heterological” one would have to restrict the quantifier in (HET) to all (one-place) predicates in the object-language. Call the restricted definition of “heterological” (HET)*.

Since “heterological” is clearly a metatheoretic predicate, it cannot be applied to (HET)*, and as a consequence no paradox gets off the ground.

3.3 *Arguments of individual self-application*

In arguments of individual self-application the substitution of an *individual constant of A* for *x* leads to a contradiction or to a counterintuitive result within the theory *T*:

> *Formal structure of arguments of individual self-application*:
> *A* is a statement (for example, a definition, an axiom, a law, a thesis) of a theory *T* and has the following form of a universal sentence:
> (*A*) For all *x* which are of kind *K*: *B*(*a*,*x*)).
> *a* is itself of kind *K*. Therefore, *B*(*a*,*a*). *B*(*a*,*a*) generates a contradiction or reveals seriously counterintuitive results of *T*. Therefore, *T* has to be rejected in its present form.
> (*B*(*a*,*x*) is a formula in which *x* occurs free, *a* is an individual constant, and *B*(*a*,*a*) results from *B*(*a*,*x*) by substituting *a* for *x* in *B*(*a*,*x*).)

For an example of an argument of individual self-application, let’s consider the so-called *Barber Paradox*:[13] A barber is ordered to shave all adult males in his village who do not shave themselves. This barber is, of course, a male adult living in this village. So, if he shaves himself, he shouldn't do it according to the order, and if he doesn't shave himself, he is one of the candidates that he should shave. We have thus the following universal claim:

> (BARB) For all adult males *x* who live in the same village as the barber *a*: *a* shaves *x* if and only if *x* does not shave *x*.

Since *a* is a male adult who lives in this village, we can substitute "*a*" for "*x*" in (BARB) and get the following contradiction:

> *a* shaves *a* if and only if *a* does not shave *a*.

The paradox generated by this argument of individual self-application can easily be dismissed. As we have seen, one of the premises of all arguments of self-application is that the universal sentence *A* is a statement *that follows* from a given theory *T*. But the order the barber gets to shave all male adult villagers who do not shave themselves clearly does not threaten any conventional theory about barbers or shaving. The sentence (BARB) does not follow from any of those theories. All that follows is that such a barber *a* cannot exist. So, the Barber Paradox is a clever riddle, but cannot be used as a devastating argument of self-application.

A similar argument of individual self-application gives rise to the famous *Russell paradox.* The *axiom of comprehension* in Frege's set-theory (see Axiom V in Frege 1893) states that every object *x* has a certain property if and only if *x* is an element of the set of all objects having that property:

> For all objects x: $A(x) \leftrightarrow x \in \{y \mid A(y)\}$. For the property of a set x not being a member of itself ($x \notin x$), this means:
> (RUSS) For all sets x: $x \notin x \leftrightarrow x \in \{y \mid y \notin y\}$.

Since the set of all sets that are not members of themselves ($\{y \mid y \notin y\}$) is, according to Frege's set-theory, itself a set, we can substitute "$\{y \mid y \notin y\}$" for "*x*" in (RUSS) and get the following contradiction:

[13] For discussions of this paradox, see, e.g., Champlin 1988: 172–174 or Rescher 2001: 143–145.

$\{y \mid y \notin y\} \notin \{y \mid y \notin y\} \leftrightarrow \{y \mid y \notin y\} \in \{y \mid y \notin y\}$.

The paradoxical result reached by this argument of individual self-application had an enormous impact on the foundations of set theory. *Prima facie* the axiom of comprehension is intuitively plausible. Furthermore, the property of a set not being a member of itself seems to be perfectly intelligible. Nevertheless, the above argument clearly shows that any set-theory in which (RUSS) can be derived is inconsistent. Consequently, if we wish to develop a consistent set-theory we must abandon some of our previously held intuitions.[14]

4. *Arguments of iterative application*

There is another type of argument which has some connection to arguments of self-application. Instead of directly applying the thesis *A* of a theory *T* to itself, as in arguments of propositional self-application, in these arguments an *instance of A* is applied to *A*, so that a certain kind of iterative usage of *A* is generated that leads to inconsistent or counterintuitive results within *T*. I will call these kinds of arguments "*arguments of iterative application*". Their general formal structure can be put as follows:

> *Formal structure of arguments of iterative application*:
> *A* is a thesis of a theory *T* and has the following form of a universal sentence:
> (*A*) For all x which are of kind *K*: $B(x)$.
> $B(x^*)$ is itself of kind *K*. Therefore, $B(B(x^*))$. $B(B(x^*))$ generates a contradiction or reveals seriously counterintuitive results of *T*. Therefore, *T* has to be rejected in its present form.
> ($B(x)$ is a formula in which x occurs free, and $B(\mathrm{x}^*)$ results from $B(x)$ by substituting the constant x^* for x in $B(x)$.

Arguments of iterative application are not self-referential in the sense that arguments of self-application are. In arguments of self-application a uni-

[14] One strategy is illustrated in Ernst Zermelo's axiomatic theory in which the axiom of comprehension is restricted in such a way that no contradiction follows. Another solution to the paradox is Russell's type-theory in which a hierarchy of sets of different types is constructed such that a formula "$x \in x$" or "$x \notin x$" is not well-formed. In both cases, (RUSS) does not follow from the given set-theory (in Russell's type-theory, (RUSS) is not even a well-formed sentence).

versal statement *A* of a theory *T* or a predicate term or a singular term of *A* falls into the domain of *A*'s universal quantifier. Whereas in arguments of iterative application, it is an *instance* of *A* that functions as a subject term of *B*(*x*). I will nevertheless briefly discuss these arguments since, as we will see, they give rise to similar problems due to the universality claim of the given theory *T*.

In what follows, I will present an argument of iterative application showing a problem for *conversational forms of contextualist accounts of knowledge*.

According to *attributor contextualism*, the truth conditions of knowledge attributions are determined by the context of the *speaker* (i.e., the knowledge *attributor* or knowledge *ascriber*). Accordingly, a sentence of the form "S knows that *p* at t" is true only if the would-be-knower S satisfies the standards for knowledge that *p* at t operant in the knowledge attributor's context. These standards are determined in part by which error possibilities are salient in the speaker's conversational context. Versions of attributor contextualism have most prominently been advocated by Stewart Cohen, David Lewis, and Keith DeRose (e.g., Cohen 1986, 1988, 1998 and 2000; DeRose 1995 and 1999, Lewis 1979 and 1996). Since the lowering and raising of the standards in these accounts is determined by conversational features, i.e. by the error possibilities that are salient to the speaker, these accounts are also sometimes called theories of *conversational contextualism* (CC). If a party to the conversation draws the speaker's attention to a *p*-falsifying error possibility or mentions a *p*-falsifying error possibility that has not yet been considered*,* the standards for knowing that *p* are raised.

One important consequence of (CC) is this: If the speaker is in an ordinary standards context in which skeptical hypotheses, like brain-in-vat hypotheses, are not salient, the epistemic subject S does not have to be able to rule out the possibility of being a brain in a vat in order to know that *p*. In these ordinary standards contexts S can thus know a lot of things about the external world, like that she has hands (provided, of course, that she has hands and that she believes it).

Let S* be a specific epistemic subject which happens to be identical to the speaker (i.e. the epistemic subject ascribes knowledge to herself) and let "c_{ord}" stands for an ordinary standards context in which brains in vat-scenarios are not salient to the speaker, we can thus state a main claim of (CC) about S* knowing that *p* in c_{ord} (KNOW_{ORD})—more formally: K(S*, *p*, c_{ord})—as follows:

> ($\mathrm{KNOW}_{\mathrm{ORD}}$) For all true propositions *p*: K(S*, *p*, c_{ord}) only if S* is in a position to rule out all *p*-falsifying error possibilities salient for S* in c_{ord}.

In ordinary standards contexts, the epistemic subject can even know that she is not a brain in a vat (provided again that she is in fact not a brain in a vat and she believes she is not a brain in a vat), since in these standards brains in vat-scenarios are not salient, and therefore the epistemic subject does not have to rule out the possibility of being a brain in a vat in order to know that she is not a brain in a vat. Let ¬BIV be the proposition of not being a brain in a vat, then "K(S*, ¬BIV, c_{ord})" is true, according to (CC) (provided ¬BIV is true and S* believes that ¬BIV). Since "K(S*, ¬BIV, c_{ord})" is a true proposition, we can substitute "K(S*, ¬BIV, c_{ord})" for "*p*" in ($\mathrm{KNOW}_{\mathrm{ORD}}$) and get (+) which, as we will see, generates an argument of iterative application:

> (+) K(S*, K(S*, ¬BIV, c_{ord}), c_{ord}) only if S* is in a position to rule out all K(S*, ¬BIV, c_{ord})-falsifying error possibilities salient for S* in c_{ord}.

Let's now make the further assumption that S* is a defender of (CC) and as a result of believing in this theory, she *comes to know* that epistemic subjects can know that they are not BIVs in c_{ord}—and in particular she comes to know that K(S*, ¬BIV, c_{ord}). But since she is a contextualist, there must be a context in which this knowledge claim is true. If she were in a high standards context in which the possibility of being a brain in a vat is salient for her, she would no longer claim to know that K(S*, ¬BIV, c_{ord}). Because she cannot rule out the salient possibility of being a brain in a vat, this possibility casts doubt on the *truth* of ¬BIV. In a high standards context, S* would therefore not claim to know that if she were in c_{ord} she would still know that ¬BIV. She would rather withdraw her knowledge that K(S*, ¬BIV, c_{ord}) in high standards context.[15]

But if she knew that K(S*, ¬BIV, c_{ord}) in low standards context c_{ord}, i.e.: K(S*, K(S*, ¬BIV, c_{ord}), c_{ord}), then with (+) we would get:

> S* is in a position to rule out all K(S*, ¬BIV, c_{ord})-falsifying error possibilities salient for S* in c_{ord}.

[15] See Brendel 2003 and 2005 for a more detailed formal analysis of this problem for conversational contextualism.

S* cannot claim to know that K(S*, ¬BIV, c_{ord}) in c_{ord}, because in the very act of contemplating K(S*, ¬BIV, c_{ord}) BIV-scenarios are brought to her attention. So, the BIV-hypothesis is a K(S*, ¬BIV, c_{ord})-falsifying error possibility salient for S* that she cannot rule out. But since BIV-scenarios are not salient for S* in c_{ord}, she cannot be in c_{ord} when claiming to know that K(S*, ¬BIV, c_{ord}). Therefore it seems, that there is no context in which a contextualist S* can claim to know a main claim of (CC), namely that K(S*, ¬BIV, c_{ord}).[16]

How devastating is this argument for conversational contextualism? The strength of this argument depends, of course, on how plausible it is that "K(S*, ¬BIV, c_{ord})" is an instance of the general claim ($KNOW_{ORD}$). A contextualist could argue that she only wants to provide a theory of the truth-conditions for knowledge claims of the *object-language*. But it seems to be a serious flaw of (CC) if a contextualist cannot claim to know a main claim of her own theory.

5. *Conclusion*

I have argued that there is not one single form of self-referential argument, but rather that there are at least two different types of self-referential arguments. *Arguments of proper self-reference* are usually employed and are most successful in contexts of formalized theories. In these arguments, an explicit self-referential structure is created by a logical equivalence of a sentence *A* with another sentence *B* in which the quotation name of *A* is an individual constant that functions as the sentence subject of *B*. This equivalence then leads to an inconsistent result within the given theory *T*. But we have seen that these arguments are only successful in reaching their argumentative goal when it can be shown that the self-referential sentence is a theorem of *T*. This, in particular, means that the sentence has to be an instance of the diagonal lemma—which in turn presupposes that the notion used in the process of diagonalization is a syntactic predicate of *T*. We have seen that whether these formal conditions are fulfilled with respect to the concept of knowledge is debatable. Even if all these conditions are fulfilled, the strength of this kind of argument still rests on the question of how devastating the result is for the main ideas of the theory in question.

[16] For a similar point of "unspeakable knowledge" in contextualist accounts, see Engel 2005: 212f.

So, for example, it does not seem to be totally devastating for a defender of the existence of an omniscient being to admit that omniscience does not mean to know *all* true sentences whatsoever.

Arguments of self-application can be divided into arguments of *propositional* self-application, arguments of *predicative* self-application and arguments of *individual* self-application. In *arguments of propositional self-application*, which are clearly the most prominent type of self-referential argument in philosophy, a claim *A* of a theory *T* falls into the domain of application of *A* itself. By applying *A* to itself, a contradictory result is obtained. In *arguments of predicative self-application* a predicate-term of *A* and in *arguments of individual self-application* an individual constant of *A* falls into the domain of application of *A*, and as a result of applying these parts of *A* to *A* itself a contradiction follows.

Most simply and most successfully arguments of self-application can be criticized by providing reasons to the effect that contrary to the first impression *A* is not a statement of *T* or does not follow from *T*. In such a case, the initial argument of self-application simply turns into a *reductio* argument showing that *A* cannot be a true sentence within *T* (if *A* is at all a syntactically well-formed sentence of *T*). But when *A* is clearly a sentence of *T* (in the examples of arguments of propositional self-application discussed above, *A* was even a main thesis of *T* or a definition of a central concept of *T*), arguments of self-application still only succeed if it can be made plausible that "*A*" (or the predicate-term or the individual constant of *T*) indeed belongs to the domain of application of *A*. If so, *T* as a universal theory has to be abandoned. However, it still has to be seen how devastating it is for the theory to give up its universal applicability by exempting *A* from the range of application of *T*.

As in the other types of self-referential arguments, the strength of *arguments of iterative application* depends on whether there are compelling reasons to think that the substitution instance that gives rise to an iterative application of a claim of the theory *T* is itself justified by *T*. If it is justified, according to *T*, the theory can no longer claim to be universally valid, and so, we must either give up the theory wholesale or revise it thoroughly in order to avoid the inconsistent result.

By exposing the logical structure of the different types of self-referential arguments and arguments of iterative application, I identified several implicit premises which might be possible targets in order to criticize those arguments. As with other types of arguments, the strength of self-referential arguments depends on whether there are compelling reasons to think that the premises on which the arguments rest are justified. In some cases self-referential arguments are in fact decisive. In particular, a theory that claims to be universally valid but is open to self-referential arguments is refuted. The question then is whether giving up the theory's totality claim leads to a complete destruction of the theory or to a reconstruction and modification in which at least some of the theory's basic ideas can be retained. At a minimum, those wishing to advance such modifications to these universal theories must motivate them and not simply offer them as *ad hoc* responses to these arguments. Whether they can do so, remains to be seen.

REFERENCES

Barke, A. (2002). *The Closure of Knowledge in Context*, mentis, Paderborn.

Boolos, G. S./Jeffrey, R. C. (1989). *Computability and Logic*, third edition, Cambridge University Press, Cambridge.

Brendel, E. (2001). 'Allwissenheit und "Offenes Philosophieren"', *Erkenntnis* 54, 7–16.

— (2003). 'Was Kontextualisten nicht wissen', *Deutsche Zeitschrift für Philosophie* 51, 1015–1032.

— (2004). 'Epistemic Closure and a Solution to the Knower Paradox', *Knowledge and Belief*, (eds.) W. Löffler/P. Weingartner, Wien, 165–173.

— (2005). 'Why Contextualists Cannot Know They Are Right: Self-Refuting Implications of Contextualism', *Acta Analytica* 20, 38–55.

Brendel, E./Jäger, C. (eds.) (2005). *Contextualisms in Epistemology*, Springer, Dordrecht (reprinted from *Erkenntnis* 61, Nos. 2–3, 2004).

Champlin, T. S. (1988). *Reflexive Paradoxes*, Routledge, London.

Cohen, S. (1986). 'Knowledge and Context', *Journal of Philosophy* 83, 574–83.

— (1988). 'How to be a Fallibilist', *Philosophical Perspectives* 2, 91–123.

— (1998). 'Contextualist Solution to Epistemological Problems: Scepticism, Gettier, and the Lottery', *Australasian Journal of Philosophy* 76, 289–306.

— (2000). 'Contextualism and Skepticism', *Philosophical Issues* 10, 94–107.

— (2004). 'Knowledge, Assertion, and Practical Reasoning', *Philosophical Issues* 14: *Epistemology*, 482–491.

DeRose, K. (1995). 'Solving the Sceptical Problem', *Philosophical Review* 104, 1–52.

— (1999). 'Contextualism: An Explanation and Defense', *Epistemology*, (eds.) J. Greco/E. Sosa, Basil Blackwell, Oxford, 187–205.

Engel, M. (2005).'What's Wrong With Contextualism, and a Noncontextualist Resolution of the Skeptical Paradox', (eds.) E. Brendel/C. Jäger, 61–89.

Frege, G. (1893). *Grundgesetze der Arithmetik*, Vol. I, Jena.

Gödel, K. (1931). 'Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I', *Monatshefte für Mathematik und Physik,* 173–198.

Hales, S. D. (1995). 'Epistemic Closure Principles', *Southern Journal of Philosophy* 33, 185–201.

Grelling, K./Nelson, L. (1908). 'Bemerkungen zu den Paradoxien von Russell und Burali-Forti', *Abhandlungen der Friesschen Schule*, vol. 2, 301–324.

Kaplan, D./Montague, R. (1960). 'A Paradox Regained', *Notre Dame Journal of Formal Logic* 1, 79–90.

Lewis, D. (1979). 'Scorekeeping in a Language Game', *Journal of Philosophical Logic* 8, 339–59.

— (1996). 'Elusive Knowledge', *Australasian Journal of Philosophy* 74, 549–67.

Plantinga, A. (1983). 'Reasons and Belief in God', *Faith and Rationality: Reason and Belief in God*, (eds.) A. Plantinga/N. Wolterstorff, University of Notre Dame Press, Notre Dame.

— (2000). *Warranted Christian Belief*, Oxford University Press, Oxford.

Rescher, N. (2001). *Paradoxes*, Open Court, Chicago and La Salle, Illinois.

Tarski, A. (1935). 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* 1, 261–405; English translation in A. Tarski: *Logic, Semantics, and Metamathematics*, Oxford University Press, Oxford, 1956.

Tetens, H. (2004). *Philosophisches Argumentieren*, Beck, München.

Wittgenstein, L. (1921). *Tractatus Logico-Philosophicus* (first pub. 1921, English translation: Routledge, London, 1961).

Yates, S. (1991). 'Self-Referential Arguments in Philosophy', *Reason Papers* 16, 133–164.

Grazer Philosophische Studien
74 (2007), 199–213.

METAPHYSICS OF SCIENCE BETWEEN METAPHYSICS AND SCIENCE

Michael ESFELD
University of Lausanne

Summary

The paper argues that metaphysics depends upon science when it comes to claims about the constitution of the real world. That thesis is illustrated by considering the examples of global supervenience, the tenseless vs. the tensed theory of time and existence, events vs. substances, and relations vs. intrinsic properties. An argument is sketched out for a metaphysics of a four-dimensional block universe whose content are events and their sequences, events consisting in physical properties instantiated at space-time points, these properties being relations rather than intrinsic properties.

1. *Introduction*

Metaphysics used to be and again is the core discipline of philosophy. In the words of Frank Jackson,

> Metaphysics … is about what there is and what it is like. But of course it is concerned not with any old shopping list of what there is and what it is like. Metaphysicians seek a comprehensive account of some subject matter—the mind, the semantic, or, most ambitiously, everything—in terms of a limited number of more or less basic notions. In doing this they are following the good example of physicists. The methodology is not that of letting a thousand flowers bloom but rather that of making do with as a meagre a diet as possible. (Jackson 1994: 25)

Metaphysics, thus conceived, seeks a comprehensive account of what there is in the real world. How do we gain knowledge about the world? When it comes to explaining how the manifest phenomena are connected, science tells us what there is in the world. As Wilfrid Sellars once put it, "in the dimension of describing and explaining the world, science is the measure of all things, of what is that it is, and of what is not that it is not" ('Empiricism and the Philosophy of Mind' (1956) in Sellars 1963: 173). There is no source of philosophical knowledge about the world independent of science. In seeking a comprehensive account of everything, metaphysics is continuous with science, going beyond particular scientific theories. The thesis of this paper is that there is a mutual dependence between science and philosophy: philosophy in the sense of metaphysics needs science to know about what there is in the real world, and science needs philosophy in the sense of epistemology when it comes to developing criteria for the interpretation of scientific theories—that is, criteria for the assessment of knowledge claims contained in scientific theories.

It is common to distinguish between the epistemology and the metaphysics of science. David Papineau, for one, draws this distinction in the following way:

> The philosophy of science can usefully be divided into two broad areas. The *epistemology* of science deals with the justification of claims to scientific knowledge. The *metaphysics* of science investigates philosophically puzzling features of the world described by science. In effect, the epistemology of science asks whether scientific theories are true, whereas the metaphysics of science considers what it would tell us about the world if they were. (Papineau 1996: 1)

Whereas the philosophy of science was dominated for decades by epistemological issues under the influence of logical empiricism and its critics, the metaphysics of science has gathered momentum in the last two decades or so. The type of metaphysics that is at issue is a revisionary in contrast to a descriptive metaphysics, to use Peter Strawson's terms (1959: introduction). The justification for a revisionary metaphysics stems from science: our best scientific theories suggest the conclusion that a number of our common sense beliefs about the constituents of the world—as analysed by what Strawson calls descriptive metaphysics—are false.

In the following, I shall sketch out four examples that illustrate the interplay between science and metaphysics—examples that show how scientific results provide a content for metaphysics, namely a content that results in a revisionary metaphysics. The examples are (1) global supervenience, (2) tenseless vs. tensed theories of time and existence, (3) events vs. substances, and (4) relations vs. intrinsic properties. In conclusion, I shall sum up the resulting view of the fundamental ingredients of the world and mention the most important open issues.

2. *Example 1: global supervenience*

When metaphysics seeks a comprehensive account of everything in terms of a limited number of basic notions, some entities have to be considered as being fundamental and all the other ones as being dependent on those fundamental entities. It is common to spell that idea out in terms of global supervenience: there is a fundamental level of the world, and everything else supervenes on what there is on that level. To quote again Frank Jackson, “Any world which is a *minimal* physical duplicate of our world is a duplicate *simpliciter* of our world” (1998: 8). A minimal physical duplicate of our world can be taken to be what one gets if one duplicates the fundamental physical level of our world. Although this is a thesis of logical supervenience, it is contingent that the actual world is a world in which supervenience holds. Hence, the thesis of global supervenience with respect to the actual world cannot be established on the basis of philosophical reflection alone, but it is based upon knowledge of our world, knowledge that originates in science.

Why should one accept global supervenience? We have physical theories at our disposal that are both fundamental and universal—namely, at the current state of physics, quantum field theory and general relativity. These theories are universal because they apply to everything that there is in the natural world. Everything is a physical system in the sense that it is subject to the laws of gravitation (general relativity) and electromagnetism (quantum field theory), among others. There are some systems to which only these theories apply. Let us call the level of these systems the fundamental level of the world. Let us furthermore assume that this level consists in the distribution of physical properties at space-time points over the whole of space-time. There is nothing smaller than a point. A field is commonly defined over the whole of space-time, with field properties being attributed to the points and regions of space-time.

By contrast, chemical or biological theories, for instance, are not fundamental. It is not the case that everything that there is in the world is a chemical system (such as a molecule) or a biological system (such as an organism). The mentioned physical theories are fundamental with respect to all the other current theories of science in the sense that these theories never need to have recourse to concepts, laws and explanations from any other theories, whereas all our other theories sometimes need to have recourse to concepts, laws and explanations from general relativity or quantum field theory. Chemical or biological theories sometimes have to invoke physical concepts and laws that are in the last resort concepts and laws of general relativity or quantum field theory—for instance, in order to explain why a given chemical or biological regularity has an exception in a particular situation. If there are laws in chemistry or biology, these are *ceteris paribus* laws, whereas the laws of fundamental physics are strict laws, admitting no exceptions.

One can sum up these considerations by putting forward a *principle of the causal, nomological and explanatory completeness of the fundamental level*: *for any fundamental physical system* p (i.e. instantiation of physical properties at a space-time point), *insofar as* p *has causes, comes under laws and admits of explanations, there are causes that are only fundamental physical causes, there are laws that are only fundamental physical laws and there are explanations that employ only concepts of fundamental physics*. This principle does not exclude that for any fundamental physical system *p*, there are other causes, laws, or explanations. But such other causes, laws, or explanations, if they exist, do not contribute anything that is not contributed by fundamental physical laws, causes, and explanations.

The principle of completeness does not occur within physics. It is a principle of the metaphysics of science, belonging to a reasonable interpretation of what our fundamental physical theories tell us about the world (for an extensive argument, see Papineau 2002: appendix). If one did not endorse that principle, an unpalatable consequence would ensue: one would be committed to saying that our fundamental physical theories are either inapplicable to some phenomena in their domain (i.e. physical properties being instantiated at space-time points), because these phenomena are covered by causes, laws and explanations of a higher-level theory instead of fundamental physical causes, laws and explanations; or, if applicable, these theories are false, because there are some phenomena in their domain for which the predictions in terms of causes, laws and explanations of these theories yield the wrong results, these phenomena being under the influence of causes, laws and explanations of a higher-level theory instead (the recent criticism of the principle of completeness that Bishop 2006 voices comes down to exploring that possibility).

To illustrate that point, consider the example of mental causes and brain states—putting in brain states for fundamental physical phenomena, and mental causes for higher-level causes that are distinct from physical causes. Assume that there are mental causes that are not neurobiological causes and that bring about neurobiological effects that are caused only by them, say produce certain brain states. In that case, as regards the brain states in question, the relevant neurobiological theory would be either inapplicable, because these states are subject to the influence of certain non-biological, mental causes; or it would be false, yielding the wrong probabilities for the occurrence of the brain states in question. If, for instance, the occurrences of a mental intention of the type raising one's right arm were distinct from neurobiological states and if these occurrences of mental intentions produced neurobiological effects that are not produced by neurobiological causes, then any mental intention of that type would regularly raise the probability for brain states of certain types to occur—a probability that would be different from the neurobiological probability that takes only biological factors into account. However, there is not the slightest evidence that the physical or neurobiological laws break down in one of these ways in some area of the brain when the mental is present. Thus, in short, as far as the principle of completeness is concerned, the argument is a philosophical one (as with any claim in the metaphysics of science)—but if the argument were not true, a consequence unacceptable for science would ensue.

The principle of completeness does not imply global supervenience. It rules out causes, laws and explanations that contribute something to fundamental physical phenomena that is not provided by fundamental physical causes, laws and explanations; but it does not exclude that there are emergent epiphenomena in our world that would not necessarily be duplicated if one created a duplicate of the fundamental physical level of our world. However, admitting such emergent epiphenomena would again lead to consequences that are unacceptable to science: such emergent epiphenomena would be such that it would in principle be impossible to find any explanation for them.

In order to strengthen that point, one can draw on another principle, namely the *principle of evolution*: *everything that there is in the real world apart from fundamental physical systems developed out of fundamental physical systems*. Given evolution, making a duplicate of the fundamental physical level of the real world would amount to duplicating cosmic evolution. If anything that there is in the real world were missing in the duplicate, everything that we know from science would lead us to expect that there also is some difference in the distribution of fundamental physical properties in that duplicate with respect to the real world—difference in some spontaneous mutation, for instance (and there is no such biochemical difference without there also being a fundamental physical difference). Consequently, such a possible world would after all not be an exact duplicate of the fundamental physical level of the real world. Completeness, if conjoined with evolution, therefore provides a strong reason for endorsing global supervenience.

Note that, according to global supervenience, what there is on the fundamental level determines everything that there is in the real world, but that the issue of determinism is a different matter. Global supervenience considers the distribution of the fundamental physical properties over the whole of space-time as the supervenience base. Determinism concerns the question of whether or not the evolution in time is deterministic. If global supervenience is valid and if the real world is deterministic, having, say, the big bang as initial condition, then any duplicate of the big bang and the laws of nature would be sufficient to amount to a duplicate of cosmic evolution. But this is a much stronger thesis than global supervenience. Global supervenience says something about worlds that are a physical duplicate of our world—that is, a duplicate of the distribution of fundamental physical properties over the *whole* of space-time—, independently of whether or not there is physical determinism. If there is quantum indeterminism and, say, two possible worlds w_1 and w_2 agree until a certain time t in the distribution of the fundamental properties and then diverge because of one radioactive atom decaying in w_1 but not in w_2, then w_2 is not a minimal physical duplicate of w_1.

3. *Example 2: the tenseless vs. the tensed theory of time and existence*

Let us have a closer look at the fundamental level, the distribution of physical properties at space-time points over the whole of space-time. There are two rival theories of time as well as of existence. According to the tensed theory of existence, existence is relative to a time in the sense that only that what is present—or only that what is present and what is past—exists. What is in the future does not exist as yet, and, according to some versions of this theory, what is past does not exist any more. The tensed theory of existence implies the tensed theory of time according to which there is a flow of time; the past, the present and the future are objective modes of time, being out there in the world. The tensed theory of time, however, does not imply the tensed theory of existence.

Opposed to the tensed theory of time is the tenseless theory of time which claims that there are only temporal relations of being earlier than, simultaneous with and later than among events, but no objective modes of past, present and future. The tenseless theory of time implies the tenseless theory of existence according to which existence is not relative to a location in time in the same way as it is not relative to a location in space: everything that there is in space and time simply exists. The tenseless theory of existence, however, does not imply the tenseless theory of time.

There are philosophical arguments in favour of both these theories of time and existence. The case can be settled by taking science into account. The relevant scientific theory is special relativity. Special relativity shows that there is no objective simultaneity. Any event—in the sense of physical properties occurring at a space-time point—that is supposed to be simultaneous with other events is so only relative to a reference frame, and there is no globally preferred reference frame. Thus, there is no objective "now"—in the same way as there is no objective "here". For any space-time point, it can be claimed that it is "present" in the same way as it can be claimed that it is "here" (see, for instance, Dorato 1995: Ch. 11 to 13, in particular 186–187, 210). The reason is that, according to special relativity, spatial as well as temporal distances between events are relative to a reference frame. Invariant with respect to the choice of a reference frame is only the four-dimensional, spatio-temporal distance between any two events (or points of space-time). That is the reason why special relativity is taken to show that space and time are united in a four-dimensional entity, space-time. General relativity—and notably its application in cosmology—goes beyond special relativity; but it does not change anything with respect to what special relativity says about the relativity of spatial and temporal distances.

Special relativity hence makes a case for the tenseless theories of time and existence. Since spatial and temporal distances are relative to a reference frame, there is no basis in the physical world for upholding a tensed theory of time or existence (see Saunders 2002). Again, this claim belongs to the metaphysics of science. It is logically possible to rescue the idea of an objective present by introducing the notion of one globally privileged reference frame. That notion does not contradict special relativity. The point is that it is entirely *ad hoc*—so that, rejecting the tenseless theories of time and existence as a claim of the metaphysics of science again has consequences that are unacceptable for science.

4. *Example 3: events vs. substances*

The tenseless theories of time and existence, based on the physics of special relativity, result in what is known as the view of the world as a block universe: the whole of four-dimensional space-time is a single block so to speak, including time; everything that there is exists at a space-time point or region. What is the content of the block universe?

In metaphysics, it is common to draw a distinction between substances and events. *Substances* persist as a whole for a certain time. They have spatial parts (unless they are atoms in a literal sense), but they do not have any temporal parts. Relying on physics, an *event* can be conceived as the physical properties instantiated at a space-time point. Continuous sequences of events are *processes*. Processes have spatial as well as temporal parts. Four-dimensional entities such as processes are commonly conceived as *perdurants*, persisting by having spatial as well as temporal parts, whereas three-dimensional entities such as substances are conceived as *endurants*, persisting as a whole for a certain time, having no temporal parts.

Common sense admits both substances and processes. A volcano, for instance, is regarded as a substance, persisting for a certain time by having no temporal parts, but only spatial parts. The eruption of a volcano, by contrast, is a process, persisting for a time by having temporal as well as spatial parts. The eruption can, for instance, be first mild and then heavy.

Events and processes cannot be dispensed with in metaphysics. Even if it were possible to conceive all events as consisting in changes of the properties of substances, there would be a dualism of substances and events qua changes in the properties of substances. However, it may be possible to do without substances, recognizing only events and processes. (There is an ambiguity in the notion of a substance: If one regards space-time points as substances, they are not substances in the sense of endurants, but four-dimensional entities that have neither spatial nor temporal parts; they are not in space and time, but they are what makes up space-time). Again, there is a philosophical dispute as to whether or not one should admit substances in addition to events. Again, science is relevant to that dispute.

If one switches from a physics of three-dimensional space to a physics of four-dimensional space-time (block universe), there no longer is any need to admit substances as the entities that are the enduring foundation of change, change being the change of properties of substances, motion being change in the location of substances. Moreover, there is no need to conceive the identity of things as the identity of substances in time, because substances do not have temporal parts.

In the metaphysics of the block universe, identity can be accounted for in terms of genidentity, that is, sequences of events that instantiate the same or similar properties. In other words, the identity of any physical object in time is explained by the fact that the object is a process whose temporal parts form a continuous sequence, exhibiting similar physical properties. As regards motion, what common sense considers as the motion of a substance through three-dimensional space is explained as a continuous sequence of space-time points or regions that possess a similar physical content (a world line). Change is different physical properties instantiated at points or regions of space-time forming a continuous sequence.

Hence, given the physics of special relativity, the arguments for a metaphysics of science that is a metaphysics of events—and processes (perdurants)—, admitting no substances (endurants) are in the first place the philosophical ones of coherence and parsimony: if one makes the step to a metaphysics of a four-dimensional block universe, it is simply not coherent to recognize three-dimensional substances among the content of the block universe. Four-dimensional events and their sequences (processes) have to be accepted anyway, and they are sufficient as the furniture of the universe (see Sider 2001 as regards the philosophical arguments).

Moreover, in recent years, arguments have been developed to the effect that admitting three-dimensional substances with spatial, but no temporal parts is not consistent with special relativity. According to special relativity, the spatial distances between points depend on a reference frame. Consequently, if one subscribes to an ontology according to which there are three-dimensional macroscopic substances, their spatial figure varies from one frame of reference to another one, because the spatial distances between the points that the substance in question occupies depend on a reference frame. If, by contrast, physical objects are four-dimensional perdurants, their figure in four-dimensional space-time is not relative to a reference frame (see Balashov 1999 as well as Hales & Johnson 2003). A further argument makes the following point: since simultaneity is relative to a reference frame, a metaphysics of enduring three-dimensional objects cannot come up with a convincing theory of the coexistence (copresence) of objects. By contrast, a metaphysics of perduring four-dimensional objects, which have temporal parts, can easily include a theory of coexistence: any two four-dimensional objects coexist if and only if they have parts that are separated by a space-like interval (see Balashov 2000 and the discussion between Gilmore 2002 and Balashov 2005). If these arguments prove sound, they put the case against three-dimensional substances on a par with the case against objective simultaneity based on special relativity.

5. *Example 4: relations vs. intrinsic properties*

Up to now, I have argued in favour of the view of the basic level of the world consisting in the distribution of fundamental physical properties at space-time points over the whole of space-time, forming continuous sequences that are processes and that can be regarded as physical objects (albeit no substances in the sense of things that do not have temporal parts). Quantum physics can be seen in the first place as adding something to this view concerning the physical properties. Whereas special and general relativity can be conceived as theories about space-time notably, quantum physics is concerned with matter.

It is often taken for granted that the fundamental physical properties, instantiated at space-time points, are intrinsic properties. Intrinsic are all and only those properties that an object has irrespective of whether or not there are other contingent objects; in brief, having or lacking an intrinsic property is independent of accompaniment or loneliness (see Langton & Lewis 1998 and for a refinement Lewis 2001). All other properties are extrinsic or relational, consisting in the object bearing certain relations to other objects.

Quantum physics is usually conceived in terms of states of physical systems. The state of a system at a time can be regarded as encapsulating the properties that the system has at that time. The most striking feature of quantum theory is that the states of several systems can be entangled. In fact, starting from the formalism of quantum theory, it is to be expected that whenever one considers a complex system that consists of two or more quantum systems, the states of these systems are entangled. Entanglement is to say that it is not the case that each of the systems has a state separately. On the contrary, only the whole, that is, the complex system composed of two or more systems, is in a precise state (called a "pure state"). Philosophers of physics therefore speak of non-separability (Howard 1989) or relational holism (Teller 1986), since entanglement consists in certain relations among quantum systems. These relations give rise to correlations that are confirmed by experiments.

These relations cannot be traced back to intrinsic properties of the physical systems in question. There are no intrinsic properties of the related quantum systems on which the relations of entanglement could supervene. Quantum physics can therefore be taken to suggest a metaphysics of relations, known also as structural realism: insofar as quantum physics is concerned, the fundamental physical properties consist in certain relations instead of being intrinsic properties (see French & Ladyman 2003 and Esfeld 2004).

Again, this is a conclusion belonging to the metaphysics of science. This conclusion could be avoided by postulating intrinsic properties in the form of hidden variables that restore separability among quantum systems (that is, properties of quantum systems that are there, but whose value we cannot know). However, since the discovery of the theorem of John Bell (1964), it is clear that one would have to pay a high metaphysical as well as physical price for admitting hidden variables of that kind. Again, if one does not accept this position in the metaphysics of science—i.e., a metaphysics of relations as fundamental physical properties, not supervening on intrinsic properties –, one faces consequences that are not acceptable to science, namely being committed to hidden variables that are intrinsic properties. (The only elaborate account of quantum physics in terms of hidden variables, Bohm's theory, does not fall within the scope of that criticism, for Bohm himself interprets his theory rather in terms of relations and holism than in terms of intrinsic properties; see Bohm & Hiley 1993).

6. *Conclusion*

Let us take stock. The examples discussed in the preceding sections suggest a view of the fundamental level of the world according to which the world is a four-dimensional block universe whose content are events and their sequences, events consisting in physical properties instantiated at space-time points, these properties being—as far as quantum physics is concerned—relations rather than intrinsic properties. This certainly is a revisionary metaphysics, rejecting a tensed view of time and existence, admitting only events and processes instead of substances (endurants) and giving priority to relations instead of intrinsic properties. Everything else there is in the world supervenes on that fundamental level in the sense of the mentioned thesis of global supervenience. The rationale for this metaphysics stems from science. The claims sketched out in the preceding sections are a reasonable interpretation of what science tells us about the world. Not endorsing them would lead to consequences that are unacceptable for science.

Nonetheless, there are a number of open issues skipped in the preceding sections:

- *the relationship between space-time and matter*: In the preceding sections, I have used the terminology of physical properties being instantiated at space-time points or regions. Are the physical properties literally properties of space-time points or regions—so that matter reduces to properties of space-time? Or is there a dualism between space-time and matter fields being inserted in space-time? This is an open issue in the philosophy of general relativity in the first place. On the one hand, the project of an outright reduction of matter to space-time failed. (This is the project of Wheeler's geometrodynamics; see Wheeler 1962 and for the acknowledgement of its failure Misner, Thorne & Wheeler 1973: § 44.3–4, in particular 1205). On the other hand, there is no clear distinction between space-time and matter in general relativity: the metric field includes spatio-temporal properties, such as the spatio-temporal distances between space-time points, as well as material properties, namely gravitation. However, resolving this issue depends not only on the philosophy of general relativity, but also on the open issue of the unification of general relativity and quantum field theory.
- *the unification of general relativity and quantum field theory*: For the time being, there are two fundamental physical theories: quantum field theory and general relativity. The relationship between these two theories is not clear. It is desirable to have one fundamental physical theory. If the project of unifying quantum field theory and general relativity succeeds, the resulting scientific theory may have important repercussions for our view of the basic level of nature. To be more precise without engaging in as yet premature speculations, quantum entanglement is independent of the spatio-temporal distance of the quantum systems whose states are entangled. This may be taken as one hint among others that space-time points are not the most fundamental level of nature. The level of physical properties instantiated at space-time points is, of course, fundamental with respect to all the other known levels—such as the levels of chemical, biological properties, etc. But there may be a quantum level that is more fundamental than the level of space-time points, space-time being somehow derived from that quantum level. If

such a view were to prove sound in the future, it would have important implications for the metaphysics of the physical world (and, perhaps, the definition of "physical" itself). Nonetheless, whatever may be the future fundamental physical theory that achieves a unified treatment of the phenomena that are currently considered by two different theories, it would be unreasonable to expect that future theory to go back behind the unification of space and time as considered by general relativity or the holism that quantum entanglement manifests. Even if we ignore as yet the content of that future theory, the metaphysical direction seems clear: events instead of enduring substances, and relations instead of intrinsic properties.

- *the micro-macro relationship*: This is what the famous measurement problem in the interpretation of quantum physics is about. That problem is still unsolved. The point at issue is the extension of quantum entanglement. Is there a physical process that leads to the dissolution of quantum entanglement so that there really are macroscopic systems having well-defined properties separately—such as cats being always either alive or dead, their states not being entangled with the states of other systems (cf. the famous thought experiment of Schrödinger's cat (Schrödinger 1935: 812)? There is a physical proposal for a further development of the formalism of quantum theory in that sense, going back to Ghirardi, Rimini & Weber (1986). But that proposal faces a number of physical problems. To my mind, this issue is rather an open physical one than a metaphysical one.

These open issues show that the metaphysics of science is an unfinished business. The metaphysics of science depends on science and its progress. In a nutshell, the metaphysics of science is as hypothetical as is science. However, since there is no source of philosophical knowledge about the constitution of the world that is independent of science, this is all that can be achieved in a metaphysics of the real world—and it is sufficient to turn the metaphysics of science into an exciting business, worth engaging in.

REFERENCES

Balashov, Y. (1999). 'Relativistic objects', *Noûs* 33, 644–662.

— (2000). 'Enduring and perduring objects in Minkowski space–time', *Philosophical Studies* 99, 129–166.

— (2005). 'Special relativity, coexistence and temporal parts: a reply to Gilmore', *Philosophical Studies* 124, 1–40.

Bell, J. S. (1964). 'On the Einstein–Podolsky–Rosen-paradox', *Physics* 1, 195–200.

Bishop, R. C. (2006). 'The hidden premiss in the causal argument for physicalism', *Analysis* 66, 44–52.

Bohm, D. & Hiley, B. (1993). *The undivided universe. An ontological interpretation of quantum theory*, Routledge, London.

Dorato, M. (1995). *Time and reality. Spacetime physics and the objectivity of temporal becoming*, Cooperativa Libraria Universitaria Editrice Bologna, Bologna.

Esfeld, M. (2004). 'Quantum entanglement and a metaphysics of relations', *Studies in History and Philosophy of Modern Physics* 35B, 601–617.

French, S. & Ladyman, J. (2003). 'Remodelling structural realism: quantum physics and the metaphysics of structure', *Synthese* 136, 31–56.

Ghirardi, G., Rimini, A. & Weber, T. (1986). 'Unified dynamics for microscopic and macroscopic systems', *Physical Review* D34, 470–491.

Gilmore, C. S. (2002). 'Balashov on special relativity and temporal parts', *Philosophical Studies* 109, 241–263.

Hales, S. D. & Johnson, T. A. (2003). 'Endurantism, perdurantism, and special relativity', *Philosophical Quarterly* 53, 524–539.

Howard, D. (1989). 'Holism, separability, and the metaphysical implications of the Bell experiments', *Philosophical consequences of quantum theory. Reflections on Bell's theorem*, (eds.) J. T. Cushing & E. McMullin, University of Notre Dame Press, Notre Dame, 224–253.

Jackson, F. (1994). 'Armchair metaphysics', *Philosophy in mind*, (eds.) J. O'Leary-Hawthorne & M. Michael, Kluwer, Dordrecht, 23–42. Reprinted in *Mind, method and conditionals. Selected essays*, F. Jackson (1998), Routledge, London, 154–176.

— (1998). *From metaphysics to ethics. A defence of conceptual analysis*, Oxford University Press, Oxford.

Langton, R. & Lewis, D. (1998). 'Defining "intrinsic"', *Philosophy and Phenomenological Research* 58, 333–345. Reprinted in *Papers in metaphysics and epistemology*, D. Lewis (1999), Cambridge University Press, Cambridge, 116–132.

Lewis, D. (2001). 'Redefining "intrinsic"', *Philosophy and Phenomenological Research* 63, 381–398.

Misner, C. W., Thorne, K. S. & Wheeler, J. A. (1973). *Gravitation*, Freeman, San Francisco:.

Papineau, D. (1996). 'Introduction', *The philosophy of science*, (ed.) D. Papineau, 1–20, Oxford University Press, Oxford.

— (2002). *Thinking about consciousness*, Oxford University Press, Oxford.

Saunders, S. (2002). 'How relativity contradicts presentism', *Time, reality & experience*, (ed.) C. Callender, Cambridge University Press, Cambridge, 277–292.

Schrödinger, E. (1935). 'Die gegenwärtige Situation in der Quantenmechanik', *Naturwissenschaften* 23, 807–812, 823–828, 844–849.

Sellars, W. (1963). *Science, perception and reality*, Routledge, London.

Sider, T. R. (2001). *Four-dimensionalism*, Clarendon Press, Oxford.

Strawson, P. F. (1959). *Individuals. An essay in descriptive metaphysics*, Routledge, London.

Teller, P. (1986). 'Relational holism and quantum mechanics', *British Journal for the Philosophy of Science* 37, 71–81.

Wheeler, J. A. (1962). *Geometrodynamics*, Academic Press, New York.

Grazer Philosophische Studien
74 (2007), 215–237.

COULD ANYTHING BE WRONG WITH ANALYTIC PHILOSOPHY?

Hans-Johann GLOCK
Universität Zürich

Summary

There is a growing feeling that analytic philosophy is in crisis. At the same time there is a widespread and *prima facie* attractive conception of analytic philosophy which implies that it equates to good philosophy. In recognition of these conflicting tendencies, my paper raises the question of whether anything could be wrong with analytic philosophy. In section 1 I indicate why analytic philosophy cannot be defined by reference to geography, topics, doctrines or even methods. This leaves open the possibility that analytic philosophy is a style of philosophizing (section 2). According to what I call a *rationalist* conception, the distinguishing feature of analytic philosophy is that it is guided by the ideal of rational argument. This conception implies that 'analytic philosophy' is an honorific title. In section 3 I point out that the rationalist definition yields a different extension for 'analytic philosophy' than commonly recognized. Section 4 defends the appeal to ordinary use in debates about the nature of analytic philosophy. Section 5 grants that there is an honorific use of the label, while also pointing out that the rationalist-cum-honorific conception is at odds with a more wide-spread and entrenched taxonomic practice. Section 6 alleges that the rationalist conception boils down to a 'persuasive definition' of analytic philosophy, and argues in favour of a more neutral philosophical taxonomy. Section 7 argues that analytic philosophy is an intellectual tradition held together *both* by lines of influence *and* by family-resemblances. The consequences for my topic are two-fold. First, there *could* obviously be something wrong with this intellectual tradition; secondly, the question whether there *is* something wrong needs to be raised separately with respect to individual phases or sections of that tradition.

Analytic philosophy is roughly 100 years old, and it is now the dominant force within Western philosophy (e.g. Searle 1996: 1–2). It has prevailed for several decades in the English-speaking world; it is in the ascendancy in Germanophone countries; and it has made significant inroads even in places once regarded as hostile, such as France. At the same time there are continuous rumours about analytic philosophy being in "crisis", or even "defunct", and complaints about its "widely perceived ills" (e.g. Biletzki/Matar 1998: xi; Leiter 2004b: 1; Preston 2004: 445–7, 463–4). A sense of crisis is palpable not just among commentators but also among some leading protagonists. von Wright noted that in the course of graduating from a revolutionary movement into the philosophical establishment, analytic philosophy has also become so diverse as to lose its distinctive profile (1993: 25). This view is echoed by countless observers who are convinced that the customary distinction between analytic and continental philosophy has become obsolete (e.g. Glendinning 2002; Bieri 2002).

Loss of identity is one general worry, loss of vitality another. Putnam has repeatedly called for "a revitalization, a renewal" of analytic philosophy (e.g. 1992: ix). And Hintikka has maintained that the survival of analytic philosophy depends on a new start based on exploiting "the constructive possibilities" in Wittgenstein's later philosophy (1998: 259, 267). Even Searle concedes that in changing from "a revolutionary minority point of view" into "the conventional, establishment point of view" analytic philosophy "has lost some of its vitality" (1996: 23; see also Williams 1996: 26). Small wonder that those hostile to analytic philosophy have for some time now been anticipating its imminent demise and replacement by a "post-analytic philosophy" (e.g. Rajchman/West 1985; Baggini/Stangroom 2002: 6; Mulhall 2002).

This combination of triumph and crisis provides a fitting opportunity to address the nature and merit of analytic philosophy from a fresh perspective. Initially I had planned to tackle the question "What, if anything, is wrong with analytic philosophy?". But where do you start? It also dawned on me that the question is based on an assumption which does not go without saying, namely that something *might* be wrong with analytic philosophy. For there is a widespread and *prima facie* attractive conception of analytic philosophy which implies that it equates to good philosophy. In recognition of this fact, the focus of this paper will be on the question: "*Could* anything be wrong with analytic philosophy?". This not only makes the subject more manageable, it also (largely) avoids the wider cultural issues that would have to be confronted in pursuit of the original question, issues which may seem out of place at a conference devoted to serious problems in metaphilosophy and epistemology.

In the first section I shall distinguish different *types* of definitions or explanations of analytic philosophy, and I shall indicate very briefly why definitions by reference to geography, topics, doctrines or even methods fail. This leaves open the possibility that analytic philosophy is a style of philosophizing. Section 2 introduces one such stylistic definition. According to a rationalist conception, the distinguishing feature of analytic philosophy is that it is guided by the ideal of rational argument. This conception leads pretty immediately to treating "analytic philosophy" as an honorific title. In section 3 I point out that the rationalist definition yields a different extension for "analytic philosophy" than commonly recognized. In section 4 I defend the appeal to ordinary use in debates about the nature of analytic philosophy. Section 5 grants that there is an honorific use of the label, while also pointing out that the rationalist-cum-honorific conception is at odds with a more wide-spread and entrenched taxonomic practice. In section 6 I allege that the rationalist conception runs the risk of boiling down to a "persuasive definition" of analytic philosophy, and I argue in favour of a neutral philosophical taxonomy. Section 7 presents my favourite conception of analytic philosophy, namely as an intellectual tradition held together *not just* by lines of influence *but also* by family-resemblances. The consequences for my general topic are two-fold. First, there could obviously be something wrong with this intellectual tradition; secondly, the question whether there is something wrong needs to be raised separately with respect to individual phases or sections of that tradition.

1. *Different approaches to the nature of analytic philosophy*

In the 1970s, Michael Dummett reopened the debate about the historical origins and the nature of analytical philosophy. At roughly the same time Anglophone academics introduced into the curriculum a subject by the name of "continental philosophy", the study of a kind of philosophy which derives its identity largely from the contrast with the analytic "Other". As a result of these two developments, there is now a booming market in different conceptions of analytic philosophy.[1] But within the often

[1] Some characterizations of analytic philosophy are clearly intended as definitions of some kind, in the sense that *ipso facto* those included do and those excluded do not qualify as analytic philosophers (e.g. Cohen 1986: ch. 2; Dummett 1993: ch. 2; Hacker 1996: 195; Føllesdal 1997). Others are formulated baldly and without qualification—"Analytic philosophy is …", "Analytic philosophers do …", "An analytic philosopher would never …" but may be intended as non-analytic generalizations, without specifying what constitutes analytic philosophy. In what follows I shall disregard this difference, and refer to both types of accounts as conceptions of analytic philosophy.

confusing variety of specific accounts, one can distinguish a few more general approaches. It is also relatively easy to see that many of them face immediate difficulties.

Geo-linguistic conceptions

In so far as analytic philosophy is contrasted with continental philosophy, it is natural and still surprisingly common to conceive of it in geo-linguistic terms. On the one side, we find what is often referred to as "Anglophone analytic philosophy", on the other side we find the kind of philosophising pursued in continental Europe.

But the analytic/continental divide is a misnomer. Taken literally, it involves "a strange cross-classification—rather as though one divided cars into front-wheel drive and Japanese" (Williams 1996: 25). Indeed, no one would think of analytic philosophy as a specifically Anglophone phenomenon, if the Nazis had not driven many of its pioneers out of central Europe. There is a variation on the Anglocentric picture, according to which analytic philosophy is at any rate Anglo-Austrian in origin and character. Proponents of the so-called "Neurath-Haller thesis" contrast an Anglo-Austrian axis of light with a Franco-German axis of darkness forever benighted by Kant. This fails not just because of imposing figures like Frege, Reichenbach and Hempel, but also because many of the most important representatives of the allegedly "Austrian" philosophical tradition were German—Brentano, Schlick, Carnap—and/or propounded a general conception of philosophy with strong Kantian affinities—Schlick, Carnap, Wittgenstein (see Glock 1997). In any event, at present analytic philosophy flourishes in many parts of the continent, while continental philosophy is highly popular in the Anglophone world, not just among literary and social theorists, but also among professional philosophers. Consequently, analytic philosophy is simply not a geo-linguistic category.

Topical conceptions

Another widespread prejudice about analytic philosophy is that it tends to concern itself with a very narrow set of topics from theoretical philosophy, in particular from logic, philosophy of language, philosophy of science and metaphysics. Many of the pioneers of analytic philosophy were indeed preoccupied with these areas. But not all of the trailblazers restricted themselves to theoretical philosophy, as the cases of Moore, Russell and Neurath testify. More importantly, at present there is literally no area of philosophy that has escaped the attentions of analytic philosophers, whether it be aesthetics, the philosophy of the body, or feminist theory. For any significant area of human thought X, there is not just a *philosophy* of X but also an *analytic* philosophy of X. Indeed, with respect to more peripheral or more recent areas such as environmental and bio-ethics, the analytic philosophy of X often preceded alternative approaches.

Doctrinal conceptions

Some commentators define analytic philosophy by reference to a particular doctrine. To be at all plausible as a unifying feature, the doctrines invoked must be suitably general and have implications concerning the aim and nature of the philosophical enterprise. Dummett's well known definition of analytic philosophy as based on the view that an analysis of thought can and must be given by an analysis of language satisfies these demands (1993: 4–5). The same holds for the mid-century view of analytic philosophy as hostile to metaphysics, for Hacker's claim that analytic philosophy is guided by the conviction that philosophy is an investigation into our conceptual scheme and hence qualitatively distinct from the special sciences (1996: 195, 319–20), and for the currently popular view that it is essentially naturalistic in outlook.

Unfortunately, all of these definitions exclude too much: Moore and Russell in the case of the linguistic and the anti-metaphysical conceptions, Russell and Quine in the case of Hacker. All three also fail to fit the current mainstream of Anglophone philosophy. The naturalistic conception, on the other hand, does more justice to that mainstream. Yet it fails for Frege, Moore, Wittgenstein and his followers, Oxford conceptual analysis, Kripke and Putnam—to name just the uncontroversial counterexamples. Furthermore, these definitions also include too much. Versions of the idea that thought should be understood through its linguistic expression would be accepted by important members of the hermeneutic and post-structuralist movements, such as Heidegger, Gadamer and Derrida. Nietzsche anticipated and indeed influenced analytic animadversions to metaphysics in Wittgenstein and logical positivism almost to the same extent as Hume. The idea of philosophy as a second-order reflection on the conceptual preconditions of first-order scientific thought was a neo-Kantian common-place. Finally, naturalists were cheaper by the dozen in the nineteenth and early twentieth century (Mill, the German physiological naturalists, psychologistic logic, Dewey, Satayana). Yet their doctrines were ripped apart by pioneers of analytic philosophy like Frege, Moore and Russell.

Methodological conceptions

The widely acknowledged failure of doctrinal definitions has encouraged definitions that are methodological. A blindingly obvious suggestion is to take seriously the "analytic" in "analytic philosophy", and to define the movement as one that pursues philosophy as analysis. Unfortunately, even *within* the context of the analytic tradition, "analysis" means diverse and often incompatible things.

One possibility (Monk 1997; see also Hacker 1997a, 56) is to understand the term "analytic" literally, namely as referring to a decomposition of complex phenomena into simpler constituents. But both the later Wittgenstein and Oxford conceptual analysis denied that propositions have ultimate components or even a definite structure. To them, analysis means the explanation of concepts and the description of conceptual connections by way of implication, presupposition and exclusion. This activity still qualifies as "connective analysis" in Strawson's sense. (1992, ch. 2). But as Strawson himself points out, the term "analysis" is misleading in so far as this procedure is no longer analogous to chemical analysis, and it might be more apposite to speak of "elucidation" instead.

The idea of a breakdown into ultimate components should also be anathema to Quineans, on account of their faith in the indeterminacy of meaning and reference. Indeed, in one respect it sits uneasily with the whole of ideal language philosophy. In that strand of analytic philosophy analysis is not the decomposition of a given complex into its components; rather, it is an act of *construction*. Thus for both Carnap and Quine analysis means "logical explication". The objective is not to provide a synonym of the *analysandum*, or even an expression with the same necessary and sufficient conditions of application. It is rather to furnish an alternative expression or construction which serves the cognitive purposes of the original equally well, while avoiding its scientific or philosophical drawbacks (Quine 1960: 224, §§33, 53–4). Finally, neither conceptual elucidation nor formal construction play a prominent role in contemporary moral philosophy and moral psychology. Frankfurt and Bernard Williams, for instance, are considered to be analytic philosophers. Nonetheless, the only sense in which they *analyse* phenomena like motivation or truth-fullness is so general, it also includes Kant's Transcendental Analytic, Nietzsche's account of morality, and perhaps even Heidegger's existential "analysis".

2. *The rationalist-cum-honorific conception of analytic philosophy*

It is tempting to think that the shortcomings of the methodological definition can be rectified by modifying it into a *stylistic* definition. What holds analytic philosophy together and distinguishes it from continental philosophy, the story goes, is not so much a specific method but a more loosely defined style of thinking and writing. In this vein, Bernard Williams assures us that what marks out analytic philosophy is "a certain way of going on which involves argument, distinctions, and, so far as it remembers to try to achieve it and succeeds, moderately plain speech". Unfortunately, the speech of many contemporary analytic philosophers is as plain as a baroque church and as clear as mud. Williams has a comeback to this objection:

> As an alternative to plain speech, it [analytic philosophy] distinguishes sharply between obscurity and technicality. It always rejects the first, but the second it sometimes finds a necessity. This feature peculiarly enrages some of its enemies. Wanting philosophy to be at once profound and accessible, they resent technicality but are comforted by obscurity (1985: vi).

Analytic philosophers for their part will no doubt find comfort in the idea that the indigestible nature of their writings is a necessity, and a sign of technical proficiency, by contrast to the wilful and whimsical obscurantism of continental authors. In fact, however, many so-called technicalities serve no purpose other than that of adopting a certain intellectual posture.

Clarity, including clarity achieved by formal devices, may have been a characteristic feature of analytic philosophy when it was dominated by writers such as Frege, Moore, Tarski, Ryle, Austin, Carnap, Reichenbach, Hempel, Quine or Strawson.[2] Even in the olden days there were notable exceptions, such as Wittgenstein, Anscombe or Sellars. And at present aspiring philosophical authors could gain more by studying continental writers like Schopenhauer, Marx, Nietzsche, or Sartre than by emulating articles in *Mind* (Glock 1998: 91–3; 2004: 432–5; see also Leiter 2004b: 11–2).

[2] Even Russell's case requires qualification. His reputation for lucidity rests mainly on works which he composed after he was forced to make a living from writing for a wider audience, and certainly not on his seminal writings between 1905 and 1910.

This leaves the second aspect of Williams' description. Perhaps not all analytic philosophers command or even aspire to a lucid style of writing. But, the story continues, they all seek a more substantial clarity. A clarity of thought rather than expression, one which involves conceptual distinctions and ultimately aims at transparency of arguments. This suggestion is captured by what I call the *rationalist* conception of analytic philosophy.[3] It holds that analytic philosophers are marked out by their rational approach to the subject, by their attempt to solve philosophical issues through argument.

In this spirit, Jonathan Cohen has referred to analytic philosophy as *The Dialogue of Reason*.[4] In a less lyrical fashion, Dagfinn Føllesdal explains analytic philosophy as a general attitude towards philosophical problems and doctrines, namely one that tackles them in a rational way, through argument.

> The answer to our question [s.l. What is analytic philosophy?] is, I believe, that analytic philosophy is very strongly concerned with argument and justification. An analytic philosopher who presents and assesses a philosophical position asks: what *reasons* are there for accepting or rejecting this position? (1997: 7).

In line with this definition, Føllesdal treats "analytic" as a scaling adjective. He classifies thinkers from very disparate schools, including apparently continental ones like phenomenology or hermeneutics, as more or less analytic depending on the role rational argument plays in their work.

Like Cohen, Føllesdal proffers this characterization with an *apologetic* intent, as part of a defence of analytic philosophy. Hence his title: 'Analytic Philosophy: what is it and why should one engage in it?'. Answering both questions, he draws the following "final conclusion":

> We should engage in analytic philosophy not just because it is *good* philosophy but also for reasons of individual and social ethics (1997: 15).

[3] A terminological remark. I shall use the term "rationalist" to include not just the continental rationalists with their emphasis on innate ideas and a priori knowledge, but any position which stresses that our beliefs should be subject to critical scrutiny and supported by argument, no matter whether these arguments invoke reason or experience. Similarly, I use to the term "reason" for the general ability to justify one's actions and beliefs by way of argument, and not in the narrow (and, in my view, misguided) sense employed by modern theories of rationality, in which it refers to a disposition to act exclusively in one's own interest.

[4] Strictly speaking, Cohen's portrayal of analytic philosophy (1986: ch. II) combines a rationalist definition with a topical approach, since he maintains not just that analytic philosophers employ reason, but also, rather implausibly, that reason is the ultimate topic of all their investigations.

A similarly uplifting spirit seems to have prevailed at the founding session of GAP, the German Society of Analytic Philosophy (at Berlin in 1990). Commenting on the proposed aims of the society, one pundit summed it all up by saying: "Perhaps we shouldn't establish a society for analytic philosophy, but simply one for *good* philosophy!".[5]

Less tongue-in-cheek, the current president of GAP, Ansgar Beckermann, explicitly and immediately connects the rationalist conception of analytic philosophy to the idea that it equates to good philosophy. According to Beckermann, "what characterizes analytic philosophy today"—after the failure of the attempt to overcome philosophy through logical analysis of language—is acceptance of two views: First, that philosophy seeks to answer substantive (rather than historical) questions in a way that is both systematic and governed by universally applicable standards of rationality; secondly, that this ambition can only be achieved if the concepts and arguments philosophers employ are made as clear and transparent as possible. "And in my view these are indeed also the distinguishing features of good philosophy" (2004: 12).

The rationalist conception has an obvious implication for our current topic. More or less explicitly, more or less intentionally, proponents of the rationalistic conception use "analytic philosophy" as an *honorific* title. Rightly so, given their assumptions. For it is surely advantageous and indeed indispensable to philosophy that it should be pursued in a rational fashion, through arguments informed by logic and conceptual distinctions.[6] Accordingly, my initial question is based on a mistaken presupposition. Nothing could be wrong with analytic philosophy, since analytic philosophy is good philosophy, granted only that there is nothing wrong with doing philosophy as such!

[5] Communication from Ansgar Beckermann, 31.08.05

[6] Even on the rationalist conception, analytic philosophy need not simply equate to good philosophy. For there are other philosophical virtues with which the unfettered pursuit of rational debate and philosophical criticism might come into conflict, for instance a concern with insights rather than argument, or with a non-aggressive academic environment. But for a rationalist, analytic philosophy is *pro tanto* good philosophy, since it satisfied an essential desideratum of sound philosophizing.

3. *The extensional problem of the rationalist conception*

The rationalist conception of analytic philosophy rightly eschews the more direct approaches discussed above. In particular, it has the advantage of allowing for the fact that analytic philosophy is a very broad church indeed. Nevertheless, it suffers from at least two shortcomings. One is that it amounts to a "persuasive definition" (see sections 5 and 6). The more immediate problem, which it shares with the aforementioned approaches, is that it is not in keeping with the commonly recognized extension of the term "analytic philosophy". In driving home this point, we need to turn the spirit of the rationalist conception against the letter, and draw a few distinctions.

The first is between theory and practice, the second between ambition and achievement. If extolling the virtue and importance of reason in theory were the decisive test, then Hegel would qualify with flying colours, notwithstanding Einstein, who compared Hegel's writings to the "drivel of a drunk". At the same time, the later Wittgenstein and some of his followers would be excluded, and so would many neo-Humeans, neo-pragmatists and sceptics. Admittedly, this unpalatable consequence might be avoided by distinguishing between

- irrationalism, a neglect of empirical science, logic, conceptual clarity and rational argument in favour of religious, political or artistic styles of thinking;
- anti-intellectualism or voluntarism, the denial that reason and intellect have the exalted position accorded to them by philosophical tradition.

A suitably revised version of rationalism has it that analytic philosophy eschews irrationalist practice, without necessarily repudiating anti-intellectualist doctrine.

At this juncture the contrast between ambition and achievement comes into play. Do you need to *succeed* at backing your claims by arguments in order to qualify as an analytic philosopher by rationalist lights? Or is it sufficient to make *bona fide* efforts? In the former case, "analytic philosopher" would be a category that can be used rarely, if ever, with any degree of confidence (a point to which I shall return below). In the latter case, the rationalist definition still faces counter-examples.

It is obvious that on this construal the rationalist definition *includes* too much. It tends to make the bulk of philosophy analytic; indeed, it bars only prophets or sages like Pascal, Kierkegaard, Nietzsche and Heidegger, and not even all parts of their work (certainly not *The Genealogy of Morals*). At least since Socrates, the attempt to tackle fundamental questions by way of reasoned argument, rather than, for example, through an appeal to authority or revelation, has been regarded as one of the features that distinguishes philosophy as such from religion or political rhetoric.

More surprisingly, it may still *exclude* too much. The early Wittgenstein is *a* paradigm case of an analytic philosopher. Not only did he initiate the linguistic turn that characterized the middle phase of analytic philosophy, he was also the first to think through the consequences of an atomist programme of logical and conceptual analysis (thereby combining previously existing Russellian and Moorean strands). And yet, he was not exactly keen to spell out the arguments behind his statements. Doing so would "spoil their beauty", he maintained in 1912, to which Russell trenchantly replied that he should acquire a slave to take over this task (Letter to Ottoline Morrell 27.5.12; quoted Monk 1996: 264).[7]

Furthermore, at least some disciples of Wittgenstein who can lay claim to being analytic philosophers, such as John McDowell, are equally immune to the ethos of the knock-down argument. Indeed, the importance of argument is played down even by some mainstream figures (e.g. Martin 2002: 133–6).

4. *In defence of ordinary use*

In criticizing the rationalist conception I have more or less explicitly appealed to the ordinary use of "analytic philosophy", its cognates and antonyms. More specifically, I have invoked its *commonly acknowledged extension*. Many contemporaries may find any such appeal outdated and downright offensive. But they should be reminded of a few points.

I. Aristotle, the first to embark on a systematic search for a conception of philosophy in general, started out from the way people used the term *sophia* (*Metaphysics* I.2; see Tugendhat 1982: ch.2).

[7] "I told him he ought not simply to *state* what he thinks true, but to give arguments for it, but he said arguments spoil its beauty, and that he would feel as if he was dirtying a flower with muddy hands. He does appeal to me—the artist in intellect is so very rare. I told him I hadn't the heart to say anything against that, and that he had better acquire a slave to state the arguments."

II. This blatant appeal to authority may merely reinforce your conviction that I am not an analytic philosophy by the lights of the rationalist conception. However, the proponents of that conception themselves rely on the recognized extension of "analytic philosophy" in their criticisms of alternatives. It is only on the assumption that a definition of "analytic philosophy" should be reportive rather than purely stipulative that commonly acknowledged analytic philosophers can be invoked as counter-examples to a proposed definition.

III. Still, it may be objected, we have only graduated to the level of a *tu-quoque* argument. The good news, however, is that both Aristotle and the rationalists are *right* to set store by the ordinary use of their respective *definienda*. In pursuing any question of the form "What is X?" we shall inevitably rely on a *preliminary notion* of X, an idea of what constitutes the topic of our investigation. As regards our case, we presuppose a preliminary understanding of analytic philosophy. This is not a fully-articulated conception emerging from the metaphilosophical debate about what analytic philosophy is, but simply an initial idea of what those debates are about. Such a pretheoretical understanding is embodied in the established use of the term "analytic philosophy". Put differently, the way we use and understand a term is not only an innocuous starting point for elucidating its meaning, it is the *only* clue we have at the outset of our investigation. In Austin's words, while "ordinary language is *not* the last word …, it *is* the *first* word" (1970: 185).

This sentiment is echoed by Quine (e.g. 1960: ch. I). In the spirit of Quine one might insist, however, that we need to move on from ordinary use towards a more specialized one based on more exacting scrutiny of the phenomena. But there are two reasons why this is not an objection to my procedure.

IV. The term "ordinary use" is ambiguous. It may refer either to the *standard* use of a term as opposed to its irregular use in whatever area it is employed, or to its *everyday* as opposed to its specialist or technical use (Ryle 1953: 301–4). Unlike "philosophy", "analytic philosophy" is a technical term used mainly by professional academics, students and intellectuals. And surely there can be nothing wrong with matching suggested definitions against the established or standard use of the experts in the relevant field.

V. One might rejoin by deploring the fact that in this instance the "experts", so-called, are philosophers rather than scientists proficient at fathoming the real nature of things. But of course there can be no question of the label “analytic philosophy” having a single *real* or *intrinsic* meaning, independently of how we explain and apply it.[8]

VI. As it stands this is no more than the superficial if incontrovertible observation that meaning is conventional in the sense that it is *arbitrary* that we use a particular sound- or inscription-pattern to mean something specific. This leaves open the possibility that analytic philosophy is a robust distinctive phenomenon, one which has an essence to be captured by a real definition rather than a nominal definition reflecting its established use. But even if the idea of real essences applies to natural kind terms in the way envisaged by Kripke and Putnam,[9] it should be clear that the taxonomic terms of philosophy have a completely different role. Nobody could seriously suggest that “analytic philosopher” applies to all and only those creatures with the same microstructure as Rudolf Carnap or Elizabeth Anscombe, let’s say, paradigmatic analytic philosopher though they are. To appreciate this it is enough to note that the term would apply equally to non-human creatures of a certain kind. Although the labels and distinctions of natural science may be capable of “carving nature at its joints”, in Plato’s striking phrase (*Phaedrus,* 265d–266a), this cannot reasonably be expected of historical labels and distinctions.

VII. Even if one accepts my general (semantic-cum-metaphilosophical) claims, one may entertain doubts about this particular case. Peter Hacker is no stranger to appeals to ordinary use. Yet he denies that the term “analytic philosophy” *has* an established use (1997, 14). Hacker is right to point out that “analytic philosophy” is a term of art, and a fairly recent one at that. It does not follow, however, that there is no established use. As pointed out above, an *established* use need not be an everyday one. In fact, what Grice and Strawson (1956) demonstrated years ago about the terms “analytic” holds equally of the term “analytic philosophy”’. Although we may lack a clear and compelling explanation, we by-and-large agree in our application of these terms. Consider the following, presumably rhetorical, question from a circular of Continuum International Publishing Group dated 21st October 2003:

[8] As Wittgenstein reminds us: “a word hasn’t got a meaning given to it, as it were, by a power independent of us, so that there could be a kind of scientific investigation into what the word *really* means. A word has the meaning someone has given to it” (1958: 28). Similarly, Davidson writes: “It’s not as though words have some wonderful thing called a meaning to which those words have somehow become attached” (1999: 41).

[9] There are notable reasons for denying this. See, e.g. Jackson 1998: ch. 3; Hanfling 2000: ch. 2; Glock 2003: ch. III.

> Are you interested in the continental philosophy of Gilles Deleuze or Theodor Adorno, or philosophy of the analytic tradition such as Friedrich Nietzsche or Mary Warnock?

There is no gainsaying the fact that while Deleuze, Adorno and Warnock are accurately classified here, Nietzsche is not. By the same token, it would obviously count against a definition of analytic philosophy if it implied that Heidegger and Lacan are analytic philosophers while Carnap and Austin are not. It would also count against a definition if it implied that Russell and Quine are analytic philosophers, while Frege and Hempel are not. Furthermore, we agree not just on what the clear cases are, but also on what count as borderline cases for various reasons, e.g. Bolzano, Whitehead, Popper, Feyerabend, and so-called neurophilosophers. Finally, the agreement is not based on a fixed list, but can be extended to an *open class* of new cases. For instance, perusal of the relevant CVs will put most of us in a position to identify clear-cut analytic and continental cases from a list of job-applicants.

5. *Honorific and descriptive uses of "Analytic Philosophy"*

A (partial) defence for the rationalist conception emerges not from denigrating established use, but from paying closer attention to it. If the rationalist definition is correct, then the concept of analytic philosophy will be what Gallie has labelled "essentially contested" (1956). Essentially contested concepts are notions like art, democracy, justice or repression. Among the features ascribed to them in the wake of Gallie, the following are pertinent to an understanding of analytic philosophy.

First, there is a pervasive practice of using these expressions in a value-laden manner, carrying strongly positive or negative connotations.

Secondly, there is no agreement on either the extension or the intension of the concept, which is to say (for our present purposes, at least), on what the concept applies to and by virtue of what properties these instances qualify.

Thirdly, disputants typically share a small core of paradigmatic exemplars and differ over which additional candidates are relevantly similar.

This final feature certainly applies to debates about the nature of analytic philosophy. And the first two will apply if the rationalist-cum-honorific conception is correct. Debates surrounding analytic philosophy would never concern the question of whether it is a good thing, at least among those philosophers who aim to pursue the subject in a rational manner. Instead they would focus on what it takes to be an analytic philosopher, and on who actually makes the grade.

Some features of the current philosophical landscape lend this suggestion a certain plausibility. From the beginning, the internal controversies over the roots and nature of analytic philosophy have been intimately linked to passionate fights for the soul and the future of analytic philosophy. Most if not all participants have tended to identify analytic philosophy with the kind of philosophy they deem proper. And this goes at least some way towards explaining the popularity of definitions with unpalatable consequences that their proponents are fully cognisant of. Thus Dummett favours the linguistic turn, and bites the bullet of defining analytic philosophy in a way that excludes Evans and Peacocke. Hacker is convinced that philosophy is a second-order conceptual investigation, and hence allows that Quine and his disciples may no longer form part of the analytic tradition proper. Some contemporary naturalists regard analytic philosophy as based on the conviction that philosophy is part of or continuous with natural science, and seem prepared to exclude Moore, Wittgenstein and Oxford conceptual analysis from the analytic club.

Even some who are generally regarded as being outside of analytic philosophy attach a certain kudos to the label. The most extreme case is a response given by the late Jacques Derrida to a paper by Adrian Moore:

> … at the beginning of your paper, when you were defining conceptual philosophy, or analytic philosophy as conceptual philosophy, I thought: well, that's what I am doing, that's exactly what I am trying to do. So: I am an analytic philosopher—a conceptual philosopher. I say this very seriously. That's why there are no fronts … I am not simply on the "continental" side. Despite a number of appearances, my "style" has something essential to do with a motivation that one also finds in analytic philosophy, in conceptual philosophy (2000: 83–84)

Surely some mistake, especially if analytic philosophy is an inherently rational pursuit. Still, it is a mistake that supports the suggestion that "analytic philosophy" is first and first foremost a coveted label, just like democracy, if sometimes on equally flimsy grounds.

Nevertheless, unlike parenthood and apple-pie, analytic philosophy is *not* something that everyone is keen to be associated with. More importantly, the refusniks include not just Nietzscheans and post-modernists, but also figures who extol rationality, at least in theory. In Germany, for instance, there are several thinkers who lay claim to the mantle of the Enlightenment tradition without purporting to be analytic philosophers, e.g. Apel, Habermas and Henrich (notwithstanding the fact that some of their Anglo-American friends present them as analytic philosophers in order to make their acquaintances look more respectable). Furthermore, counter-instances include not just representatives of "Old Europe", but also figures in Anglophone philosophy. Here are just a few examples taken from very diverse quarters.

His intellectual proximity and debt to the Vienna Circle notwithstanding, Karl Popper distanced himself from analytic philosophy, since he regarded it as the brain-child of Wittgenstein's "no nonsense" campaign against metaphysics, and hence as devil's work. Simon Critchley (2001), an eminent expositor of continental philosophy, shuns both analytic philosophy, which he disparages as "scientistic", and the "obscurantism" of religion and New Age thinking. Jerry Fodor vehemently disavows being an analytic philosopher (see Leiter 2004b). Admittedly, his grounds for doing so are feeble. For he makes being an analytic philosophy depend not just on the linguistic turn—aka "semantic ascent"—but also on subscribing to an even more specific and less popular doctrine—"semantic pragmatism"—which explains intensional content as "some sort of 'know how'". The crucial point in our context, however, is that Fodor is happy to renounce analytic philosophy as he defines it.

Analytic philosophy may be a *contested* concept among some philosophers and within certain debates. But it is *not* an essentially contested concept. The most fundamental feature of its intension is not that it refers to an intellectual activity that is to be commended—whatever it may look like. While there is an honorific use of "analytic philosophy", the descriptive use is more widely spread and more firmly entrenched. What is more, the descriptive use provides the basis for the honorific one. People associate analytic philosophy with certain features, and they then evaluate these features differently. Understanding the term "analytic philosophy" is tied to the ability to specify certain figures, movements, texts and institutions, and perhaps some of their prominent qualities. It does not require the belief that analytic philosophy is at any rate a jolly good thing.

6. *In defence of neutrality*

Although a definition of analytic philosophy must be nominal rather than real, it is *not* a free for all. Within nominal definitions, we can distinguish between reportive or lexical definitions on the one hand, revisionary definitions on the other. The clear desideratum for reportive definitions is that they should conform to established usage and institutional practice as regards both the extension of the definiendum and its intension. But of course there may be reasons for *redefining* a term. Revisionary (nominal) definitions can in turn be divided into stipulative definitions and explications.

Stipulative definitions simply lay down *ab novo* what an expression is to mean in a particular context, in complete disregard of any established use it may have. Such definitions cannot be correct or incorrect. But they can be more or less fruitful, in that it may be more or less helpful to single out a particular phenomenon through a separate label. With respect to established terms sheer unrestricted stipulation is rarely advisable, because it invites confusion for no apparent gain. At the very least, stipulative use must always remain mindful of the established one, otherwise equivocation is guaranteed.

As we have seen, the rationalist definition is not purely stipulative. It purports to pay heed to paradigmatic instances, and its proponents would not accept, I take it, if analytic philosophy were defined as anything other than a kind of philosophy. Furthermore, it captures—more or less—one existing way of using the label "analytic philosophy". One might argue, therefore, that this honorific use is *superior* to the descriptive one. By this token, the rationalist definition would provide a kind of logical explication, a definition which avoids shortcomings of the standard descriptive use.

Yet the only potential shortcoming of that use that we have encountered so far is that it is vague, in the sense of allowing for borderline cases. This would mean that an explication should take on the form of a so-called "precising definition". It is a moot question whether vagueness is indeed undesirable in the area of philosophical-cum-historical taxonomy. But let us grant, for the sake of argument, that a precising definition avoiding vagueness is called for. Even then the rationalist definition is not an option. For instead of tidying up the rough edges of the descriptive employment of "analytic philosophy", it yields an *entirely different* extension, one that reaches way back to the 6th century BC and includes figures that are standardly classified in completely different terms.

In fact, the boot is on the other foot. Rather than solving problems, the honorific use creates new ones. The danger of cross-classification can be avoided, to be sure, if "analytic philosophy" is consistently employed as a taxonomic label of a different kind to other label, whether these be historical—e.g. scholasticism or German idealism, or doctrinal, such as Platonism or naturalism (this is an advantage of Føllesdal's version of the rationalist conception). But the honorific use still has drawbacks as compared to its descriptive rival.

The first is that the honorific label is either *too undiscriminating* or *too demanding*. Remember Fodor's disavowal of being an analytic philosopher by his (rather narrow) doctrinal definition. When pressed further by Leiter he writes:

> Oh, well, there's an uninteresting notion of "analytic philosopher" which just means "philosopher who tries to argue for his claims". I am, or at least hope some day to be, an analytic philosopher in THAT sense (Leiter 2004b).

I am proud to announce that I am one step ahead of Fodor here. I don't just *hope* to try to argue for my claims some day, I already *do* try to argue for them. My hope is that some day I shall succeed (though it probably won't be today). So I already satisfy the notion of "analytic philosopher" that Fodor describes as uninteresting. He is right to do so. For my achievement is rather minimal. As mentioned above, most if not all philosophers have *tried* to argue in some way or other for their claims. But a classification which implies that all or most philosophers qualify as analytic does less work than one which draws a line between significantly different phenomena.

But if we turn from ambition to achievement, the honorific label once more causes trouble. If soundness or even validity is required, the label will be way too demanding. For its application would presuppose accreditation of an achievement which is notoriously and, it appears, incurably contested between philosophers. Many (though by no means all) analytic philosophers worship at the shrine of the knock-down argument. But these philosophers are no closer to a consensus on what constitutes a sound argument than other philosophers.

The alternative is to allow for a category of *genuinely arguing* rather than *merely trying* which does not presuppose that the argument is compelling. Even if this category could be reasonably well defined, however, it would still imply a substantial achievement. This consequence militates against an important desideratum of philosophical taxonomy. It should be possible to classify someone as an analytic philosophy without having to decide whether he or she is a good philosopher, or at least good enough to present something that looks like it might be a compelling argument.

This problem is intimately connected to a second worry. To put it bluntly: just as theists should not be allowed to define God into existence, analytic philosophers should not be allowed to define themselves into excellence!

Of course proponents of the rationalists definition will disavow any such underhand scheme. Remaining faithful to their rationalist aspirations, they would have to grant that all bets are off. The question of who qualifies as an analytic philosopher would have to be decided afresh, without any preconceptions stemming from the descriptive use. Alas, this is easier said than done! Consider in particular the one arena in which the honorific use plays its greatest role, namely the notoriously acrimonious and ill-tempered exchanges with the despised "continentals". In this context it is particularly tempting to move from one's uncontentious membership of an intellectual tradition to avowals of intellectual superiority. Thus Searle is reported to have responded to an unfortunate interlocutor who introduced himself as a phenomenologist, "I am an analytic philosopher. I think for myself" (Mulligan 2003: 267).

In such contexts the rationalist conception clearly amounts to a "persuasive definition", one which appeals to certain preconceptions of the party to whom it is given in order to make a claim or position more persuasive. One example is to define politicians as "self-serving manipulators" in a debate about whether all politicians are immoral. The definition clearly prejudges the issue, since to manipulate others for one's own purposes is (*pro tanto*) immoral. Similarly, defining analytic philosophers as "philosophers who pursue their subject in a rational manner" prejudges the issues if one is debating the merits, or otherwise, of analytic philosophy and its rivals.

The only way of avoiding this "persuasive" abuse of the honorific label is to keep it out of certain debates. Yet this is a drawback in its own right. Definitions should foreclose as few substantive and interesting questions or debates as possible. And among these are indisputably certain questions about analytic philosophy as identified by the standard descriptive use: Is analytic philosophy good philosophy? Has it made significant advances over its predecessors? Is it superior to its current rivals? Is it making progress or at least going in the right direction?

Consequently one of the objections to the rationalist definition, though by no means the only one, is precisely that it prejudges the question of whether anything is wrong with analytic philosophy. But what precisely is the alternative? We can make a start with an observation by Sluga:

> Following common practice, I take analytic philosophy here as originating in the work of Frege, Russell, Moore, and Wittgenstein, as encompassing the logical empiricism of the Vienna Circle, English ordinary language philosophy of the post-war period, American mainstream philosophy of recent decades, as well as their worldwide affiliates and descendents (1997: 16n).

As already indicated, I think that Sluga is right to assume that there is a common practice. Moreover, I think that his list indeed conforms to that practice in its extension.[10] Furthermore, Sluga also gives a hint of what holds that tradition together. He explains analytic philosophy genetically, as a historical sequence of individuals and schools that influenced, and engaged in debate with, each other, without sharing any single doctrine, problem or method (a line also taken in Hacker 1997).

I am sympathetic to this historical conception of analytic philosophy. But I think that it is important to preserve a kernel of truth in methodological and stylistic conceptions, including the rationalist conception as developed by Føllesdal. Philosophers that do not crop up on the standard lists may have greater or lesser affinities with analytic philosophy, and they may rank among its precursors. Such claims have been made, for example, on behalf of Aristotle, Aquinas, Descartes, Leibniz, the British empiricists, Bolzano, Brentano, Husserl, and the Kantian tradition.

Obviously, to count as a precursor of analytic philosophy it is not enough to have influenced individual analytic philosophers; otherwise one would have to include, e.g., Hegel (on account of Brandom or McDowell), Schopenhauer (on account of Wittgenstein), Nietzsche (on account of Danto and Williams), and Marx (on account of Neurath and Jerry Cohen). What one needs are general features, whether thematic, doctrinal, methodological or stylistic, which unite any additions to the core list. For this reason I am in favour of combining a historical approach with yet another approach. This approach treats analytic philosophy as a

[10] By contrast to Hylton (1990, 14) whose historical conception is too narrow, excluding Moore, Wittgenstein, and Oxford philosophy: "In speaking of analytic philosophy here I have in mind that tradition which looks for inspiration to the works of Frege, Russell and Carnap".

family resemblance concept in Wittgenstein's sense. What holds analytic philosophers together is neither mere historical influence nor a single set of necessary and sufficient conditions, but a thread of overlapping similarities. Thus current analytic philosophy is tied to Frege and Russell in its logical methods, to the logical positivists in its respect for and interest in science, to Wittgenstein and conceptual analysis in its concern with the a priori and its elucidation of concepts, etc.

This provides us with a good platform for raising the issue of what, if anything is wrong with analytic philosophy. On the one hand, it identifies a group of thinkers, works, schools or institutions the value of which can be assessed. On the other hand, the fact that these are held together not just by ties of influence but also by overlapping features with substantive philosophical significance allows us to make that question more specific, by homing in on particular doctrines, methods, etc. The proper reaction to my initial question "What, if anything, is wrong with analytic philosophy?" is not a straightforward list of gripes and gongs. Nor is it the rejection of the question on the grounds that analytic philosophy is rational and *ipso facto* wholesome philosophy. It is rather the counter question: Which part of the analytic family do you have in mind?

REFERENCES

Austin, J. L. (1970). *Philosophical Papers*, Oxford University Press.

Baggini, J., J. Strangroom (eds.) (2000). *New British Philosophy: the Interviews*, Routledge, London.

Beckermann, Ansgar (2004). 'Einleitung', *Grundbegriffe der Analytischen Philosophie*, (ed.) P. Prechtl, Metzler, Stuttgart.

Bieri, P. (2002). 'Was bleibt von der Analytischen Philosophie, wenn die Dogmen gefallen sind?', CD-Rom, Einstein Forum, Potsdam.

Biletzki, A., A. Matar (eds.) (1998). *The Story of Analytic Philosophy*, Routledge, London.

Bunnin, N., E. P. Tsui-James (eds.) (1996). *The Blackwell Companion to Philosophy*, Blackwell, Oxford.

Cohen, J. L. *The Dialogue of Reason*, Clarendon, Oxford.

Critchley, S. (2001). *Continental Philosophy: a very short Introduction*, Oxford University Press.

Davidson, D. (1994) 'Intellectual Autobiography', *The Philosophy of Davidson*, (ed.) Hahn, Open Court, La Salle, 3–79.

Derrida, Jacques (2000). 'Response to Moore', *Arguing with Derrida*, (ed.) S. Glendinning, Blackwell, Oxford, 381–6.

Dummett, M. A. E. (1993). *The Origins of Analytic Philosophy*, Duckworth, London.

Fodor, J. (2004). 'What is "Analytic" Philosophy? Thoughts from Fodor', http://leiterreports.typepad.com/blog/2004/10/what_is_analyti.html accessed 27.10.04.

Føllesdal, D. (1997). 'Analytic Philosophy: what is it, and why should one engage in it?', (ed.) Glock, 1–16.

Gallie, W. B. (1956). 'Essentially Contested Concepts', *Proceedings of the Aristotelian Society* 56: 167–98.

Glendinning, S. (2000). 'The Analytic and the Continental', (eds.) Baggini / Stangroom, 201–18.

Glock, H. J. (1997). 'Philosophy, Thought and Language', *Thought and Language*, (ed.) J. Preston, Cambridge University Press, Cambridge, 151–69.

— (1998). 'Insignificant Others: the mutual Prejudices of Anglophone and Germanophone Philosophers', *Cultural Negotiations*, (eds.) C. Brown & T. Seidel, Francke Verlag, Tübingen, 83–98.

— (2003). *Quine and Davidson on Language, Thought and Reality*, Cambridge University Press.

— (2004). 'Was Wittgenstein an Analytic Philosopher?', *Metaphilosophy* 35.4, 419–44.

— (1997). (ed.) *The Rise of Analytic Philosophy*, Blackwell, Oxford.

Grice, H. P., P. F. Strawson (1956). 'In Defense of a Dogma', *Philosophical Review* (65), 141–58 .

Hacker, P. M. S. (1996). *Wittgenstein's Place in Twentieth Century Analytic Philosophy*, Blackwell, Oxford.

— (1997). 'Analytic Philosophy: What, Whence and Whither', (eds.) Biletzki/Matar, 1–34.

— (1997a). 'The Rise of Twentieth Century Analytic Philosophy', (ed.) Glock, 51–76.

Hanfling, O. *Philosophy and Ordinary Language*, Routledge, London.

Hintikka, J. (1998). 'Who is About to Kill Analytic Philosophy?', (eds.) Biletzki/Matar .

Hylton, P. (1990). *Russell, Idealism, and the Emergence of Analytic Philosophy*, Clarendon Press, Oxford.

Jackson, F. (1998). *From Metaphysics to Ethics*, Oxford University Press.

Leiter, B. (2004a) 'What is "Analytic" Philosophy? Thoughts from Fodor', The Leiter Reports: Editorials, News, Updates October 21, 2004, http://webapp.utexas.edu/blogs/archives/bleiter/002261.htm.

— (2004b). 'Introduction', *The Future for Philosophy*, (ed.) B. Leiter, Oxford University Press, 1–23.

Martin, M. (2000). 'The Concerns of Analytic Philosophy', (eds.) Baggini/Stangroom, 129–46.

Monk, R. (1996). *Bertrand Russell: the Spirit of Solitude*, Free Press, New York.

— (1997). 'Was Russell an Analytic Philosopher?', *The Rise of Analytic Philosophy*, (ed.) H.J. Glock, 35–50.

Mulhall, S. (2000). 'Post-Analytic Philosophy', (eds.) Baggini/Stangroom, 237–52.

Mulligan, K. (2003). 'Searle, Derrida, and the Ends of Phenomenology', *John Searle*, (ed.) B. Smith, Cambridge University Press, 261–86.

Preston, A. (2004). 'Prolegomena to any Future History of Analytic Philosophy', *Metaphilosophy* 35, 445–465.

Putnam, H. (1992). *Renewing Philosophy*, Harvard University Press.

Quine, W.V. (1960). *Word and Object*, MIT Press.

Rajchman, J., C. West (eds.) (1985). *Post-Analytic Philosophy*, Columbia University Press, New York.

Ryle, G. (1953). 'Ordinary Language', *Collected Essays* Vol. II, Hutchinson, London, 1971, 301–18.

Searle, J. (1996). 'Contemporary Philosophy in the United States', (eds.) Bunnin/Tsui-James, 1–24.

Sluga, H. (1997). 'Frege on Meaning', (ed.) Glock, 17–34.

Strawson, P.F. (1992). *Analysis and Metaphysics*, Oxford University Press, Oxford.

Tugendhat, E. (1982). *Traditional and Analytical Philosophy*, trans. P. Garner. Cambridge University Press, Cambridge, German edn. 1976.

Williams, B. (1985). *Ethics and the Limits of Philosophy*, Fontana, London.

— (1996). 'Contemporary Philosophy—a Second Look', (eds.) Bunnin/Tsui-James, 25–37.

Wittgenstein, L. (1958). *Blue and Brown Books*, Blackwell, Oxford.

von Wright, G.H. (1993). *The Tree of Knowledge*, Brill, Leiden.

Grazer Philosophische Studien
74 (2007), 239–250.

ON NOT FORGETTING THE EPISTEMOLOGY OF NAMES

Frank JACKSON
The Australian National University

Summary

This paper argues that the path to knowledge concerning the right account of proper names attends to their representational and epistemological roles—to, that is, their contribution in sentences of the form "*A* is *F*" to how things are being represented to be by the sentence, to the information about how things are that such sentences deliver to us, and to the way this information is used to justify the production of such sentences. These considerations, I argue, support a version of the description theory of reference for names.

I.

The philosophy of language changed forever with the publication of *Naming and Necessity* (Kripke 1980) and 'The Meaning of "Meaning"' (Putnam 1975). We can all agree about that. But there is a lively debate over precisely what we learnt from those seminal works and the literature they spawned. This paper is concerned with one part of the debate, the part most particularly concerned with the description theory of reference for names. I belong to the party that holds that the description theory of reference was transformed but not eliminated: the party the opposition think of as failing to see how deep the Kripke-Putnam revolution cuts into traditional views of meaning and reference. In particular, I am one of the minority (?) who hold that we do not learn that the reference of names does not go by descriptions but rather that it does not go by the descriptions that first come to mind when the name is mentioned, and, in addition, that we learn that rigidification is rife and that anchoring or centering is rife.

How might one reply to those who belong to the surgery school? One strategy is to offer rebuttals of each of the arguments against the description theory of reference for names. I have done a bit of that in the past

(Jackson 1998, 2003, 2005). Another, more constructive reply describes a way of thinking about reference for names that supports the description theory. This paper is of the second kind. I will say almost nothing about the famous arguments against the description theory. I want to paint, in the broad, another way of looking at the debate over the reference of names, a way that draws on some epistemological considerations.

II.

The description theory—as I will understand it, and this way of understanding it is part of the picture I want to paint—holds that sentences of the form "*A* is *F*", where "*A*" is a name, not necessarily a proper name though we will mainly discuss cases where it is a proper name, is by virtue of an implicitly agreed convention[1] a sentence to use to express a user's belief that the *D* is *F*. Statements of the description theory sometimes say that it is the view that reference for names goes by *associated* description. In these terms, "*D*" is the associated description, and it gets to be the associated description because the sentence "*A* is *F*" is a sentence for the user to use to express his or her belief that the *D* is *F*.

Some points to note by way of stage setting. i) The term "description" is a misnomer. When I believe that the *D* is *F*, I may or may not have the word "*D*" in my language. The theory should probably have been called the property theory of reference. I will use "property" and "description" interchangeably in what follows, where "property" means any pattern in nature. Joints in nature, very natural patterns and all that, are not to the point here. ii) It is no part of the description theory that everyone who understands English (say) expresses the same belief with the sentence "*A* is *F*". The associated description for a name may, and sometimes does, vary from one English speaker to another. iii) The account of "associated description (property)", given above in terms of the belief expressed, separates it sharply from the sometime sense of a property widely believed to be possessed by the referent. Most of us believe that the fattest person in the world will die very soon, but the belief that the fattest person is in the next room does not imply the belief that a person who will die very soon is in the next room. iv) Statements of the description theory often

[1] Concerning which see Grice 1957, Lewis 1969, Bennett 1976, Locke 1690: Book III, Ch. II, § 2.

talk of a cluster of descriptions for some given name. We will talk in the singular, thinking where necessary of a single property that is a disjunction of conjunctions. Nothing hangs on this. v) The reason for the slightly awkward locution "A sentence to use to express a user's belief that the *D* is *F*" is of course that there may be more than one such sentence, and the person may not have the belief in question or may not make it public in assertion. vi) It is relatively noncontroversial that reference supervenes on nature. If "*A*" refers to *x* but not to *y*, there must be some difference between *x* and *y* in regard to their properties—otherwise *x* rather than *y* being "*A*"'s referent would be *ungrounded*.[2] This means that in one sense it is relatively uncontroversial that reference goes by the distribution of properties (descriptions). But it does not follow from this that reference goes by *associated* properties, and that is what the argy bargy is about.

III.

We can capture the essence of the description theory using the notion of representation, trading on the fact that belief is the representational state *par excellence*. If "*A* is *F*" is a good sentence for me to express the belief that the *D* is *F*, then "*A* is *F*" represents that the *D* is *F* (in my mouth: in someone else's it may represent that the *E* is *F*). But a representational take on the debate is very plausible independently of the description theory of reference for names. Sentences we understand provide putative information on how things are in the way typical of systems of representation like maps and diagrams. Maps and diagrams which we understand provide putative information about how things are—the layout of the London Underground, as it might be—by virtue of there being known functions from the maps and diagrams to various ways things might be, with the putative information provided being the value of the relevant function at the way the map or diagram is: *mutatis mutandis* for sentences. (There is more on this theme in § V below.) If we think in the familiar possible worlds way, we can say it thus: the putative information provided by a map or diagram is the set of worlds where things are as the map or diagram represents them to be: *mutatis mutandis* for sentences. The description

[2] Examples involving worlds with duplicate regions tell us that there are important niceties of formulation here that would need to be included in a fuller statement; see Jackson 2003: § 3.3.

theory of reference for names is then the view that the representational content of "*A* is *F*" is the set of worlds where the *D* is *F*.

We are not here begging the question in favour of the description theory.[3] Direct reference theorists, for example, can insist that the value of the function at "*A* is *F*" is the set of worlds where *A* itself, however propertied, is *F*, and not the answer favoured by description theorists, the set where the *D* is *F*.[4]

IV.

There are two immediate implications of looking at the debate over the reference of names in terms of representation.

If most sentences and words are part of a system of representation (the exceptions will include expressions of attitudes like "Hooray" and, on some views, indicative conditionals and ethical sentences), a principal task of a theory of reference is disinter the contribution of one or another word or phrase to the representational content of some whole in which it appears. The sentence "Paris is pretty" represents that things are a certain way. A question for the theory of reference is, What is the contribution of "Paris" in this sentence to the representational content of the sentence? The answer from the description theory is that the contribution is that the sentence represents that the *D* is pretty, where *D* is the property associated with "Paris".

This is quite different from saying that "Paris" is synonymous, in anything like the traditional sense, with "the *D*", and in fact it clearly isn't. For example, as everyone agrees, "Paris" is rigid and "the *D*" typically is not. But that point is consistent with their agreeing in their contribution to the representational content of sentences of the form "—is pretty". Consider the pair: "The actual *F* is *G*" and "The *F* is *G*". They agree in how things are being represented to be. For how must the world we live in be if the first sentence is to be true? What needs to be the case is that there is one *F* and it is *G*, and exactly the same goes for the second sentence. The two sentences represent alike. But of course the "the actual *F*" is rigid whereas the "the *F*" is not. The phrases are not synonymous.

[3] And Soames 2005: 7, a strong critic of the description theory, says that it is "undeniable" that language is representational.

[4] We will represent ways things might be in the familiar possible worlds manner.

There is a general point to be made here. The fact that we finite creatures are able to grasp the representational contents of a vast range of sentences tells us that there exist systematic, graspable connections between the way parts of sentences and their organisation determine how things are being represented to be. We do not use brute force. It follows that, as a rule, a proper name "*A*"'s contribution in "*A* is *F*" bears some non-accidental relationship to its contribution in, say, "*S* believes that *A* is *F*" and in "Had things been thus and so, *A* would have been such and such" etc. But it should not be assumed that if "*A* is *F*" represents that the *D* is *F*, then "*S* believes that *A* is *F*" must represent that *S* believes that the *D* is *F*, and "Had things been thus and so, *A* would have been so and so" represents that had things been thus and so, the *D* would have been so and so. Natural languages evolved; they were not designed by logicians. A good deal of messiness is to be expected. The example of phonics is an example where there is a good deal of messiness. There are systematic, graspable connections between the way parts of words and their organisation (and sometimes their context) determine correct pronunciation, and if there weren't we could not have learnt how to pronounce English words correctly, but there is a lot of messiness.

V.

The second implication of the representational approach is that, in a sense to be explained in a moment, linguistic reference is not hidden.

Suppose I believe that the next President of the United States will be a Democrat. I have no idea who that person will be. My belief is based solely on what I see as the troubles of the current Republican administration. I express my belief using the sentence "The next President of the United States will be a Democrat". Something about reference is hidden. I do not know the reference of "the next President of the United States". However, it does not follow that I do not know how things have to be in order for my belief to be true, and in fact I do know how things have to be for the belief and sentence to be true. (And if *I* didn't, how could I expect my hearers to know—in which case why bother to produce the sentence. It could serve no useful informational role.[5]) In that sense—call

[5] I labour the point as, in discussion of these issues, I find that some seem to be happy to accept that they do not know what they are saying about how things are much of the time.

it the representational sense—I know the reference of the sentence. The representational picture of language supports the contention that those competent with sentences of the form "*A* is *F*" do know their representational reference—or, more exactly, the representational picture *plus* the manifest utility of such sentences for finding our way around our world does the supporting. Let's spell this last point out.

Maps are very useful for finding our way around the world because of *two* facts: one, there is a function from ways the map might be to ways the world might be; two, we know what that function is. (Well, there are many functions; what matters is that we know the right, intended one for a given map.) The utility requires both: maps made according to unknown conventions are useless. But as well as requiring both, it supports both. The best explanation of the utility of maps is that the many who find them useful know the relevant function. The same general line of thought tells us that, in addition to there being a function from ways sentences might be to ways the world might be, we know what that function is for any language that assists us to find our way around our world. Confrontation with sentences in English containing proper names confers on English users all sorts of abilities. Good use of the sentence "Erfurt is where the conference is" makes it possible for a group of initially scattered philosophers to end up in the same place. Names on letters help checks and speeding fines to arrive at the right destination. Names in phone books are enabling; and sentences containing names are vital in hospitals to enure that there is the right match between patent and operation. The name "Gödel" was very useful for those who wished to meet Gödel. And so on and so forth. The obvious explanation for the utility of names is that we know the function from sentences containing names to how things are being represented to be by those sentences.

Some object that this is only one explanation: another is that we are dealing here with sets of abilities that are *mere* abilities: the cases are cases of knowledge how, not knowledge that. This is a kind of instrumentalism in the philosophy of language which should, I think, be rejected for the same reason we reject it in the philosophy of science. There is much to say on this issue[6] but here I will simply note that surely it is a commonplace that the *explanation* of why maps are very useful instruments for finding one's way around is our grasp of the relevant functions from maps to how things are represented to be by maps. Looking at maps confers navigational

[6] For more, see Jackson 2005.

abilities *because* we grasp how they represent things to be. I think we should say the same about sentences, including those containing names.

VI.

Now for the epistemology.

Fred attends a lecture on Australian history and is told the following. There was someone called “Matthew Flinders”. He was born in England, was the first person to circumnavigate Australia, and he suggested the name “Australia” for Australia. Fred reflects on what he has been told, on his knowledge of when Australia was discovered and the methods of travel available at that time, and expresses his conclusions in sentences like “Matthew Flinders is dead” and “Matthew Flinders went by boat from England to Australia”. This would be warranted—it is typical of the kind of case where we use a sentence of the form “*A* is *F*” the first time we hear of *A*. But what exactly is warranted? Surely nothing more than something like: that there was someone who falls under certain descriptions—born in England, called “Matthew Flinders” and so forth—who is now dead and went by boat from England to Australia. But if a proper way for Fred to express what he has learnt is in the sentences containing the proper name “Matthew Flinders” of a kind with those just given, what else can these sentences be saying about how things are than that there was someone who falls under the descriptions who was thus and so? To say that Fred acquired more warranted opinion than this would be giving words *per se* epistemic power, and using a proper name does not in itself cut any epistemic ice.

What happens as time passes and Fred learns more and more about Matthew Flinders? Does this change matters in any essentials? It is hard to see that it does, or how it could. The world confronts us and our informants as a complex array of property instances and we produce sentences of the form “*A* is *F*” in response to putative information of that kind. We do not get information as such on primitive thisnesses and thatnesses if such there be. It is all one property instance after another. I am not saying of course that we never get information about the holding of identities. I am saying that such information is downstream from information about property instances. When astronomers showed that Hesperus is Phosphorus, they did so by establishing facts about the distribution of properties. They started with knowledge of the form:

$$\exists x\,[\mathrm{H}x \;\&\; \forall y\,(\mathrm{H}y \supset x = y)] \;\&\; \exists x\,[\mathrm{P}x \;\&\; \forall y(\mathrm{P}y \supset x = y)]$$

and proceeded to establish something of the form:

$$\exists x\,[(\mathrm{H}x \;\&\; \mathrm{P}x) \;\&\; \forall y\,[(\mathrm{H}y \;\&\; \mathrm{P}y) \supset x = y)].$$

Or suppose you had to construct a machine that, when confronted with an object, rang a bell if and only if the item in front of it, that very item, had been in front of it previously. You could only do so by utilising the effects of the items placed before the machine on the machine—the "traces" of the items—and traces are the causal product of property instances. The moral is that the only way to understand sentences of the form "*A* is *F*" so as to make their production epistemically kosher is to understand them as representing that the *D* is *F*.

The same line of thought can be applied to names of natural kinds. It is relatively (*relatively*) easy to acquire the belief that the sea contains a natural kind that is also found in lakes, falls from the sky, is potable, colourless and so on. It is also relatively easy to tag this kind "water". (In so tagging it we are not of course assuming that the kind always has the properties it has when we come across it in lakes etc.) The situation we have just described is roughly how things were before Lavoisier. We recorded how we took things to be in sentences like "The sea contains water" and we were justified in holding that things were indeed that way. However we were not justified in believing that the sea contained H_2O. That took hard work by Lavoisier. What then is the right thing to say about what we were justified in representing about how things were, before the rise of modern chemistry, with the sentence "The sea contains water"? Only something like that the sea contains the watery stuff.[7] Coining the word "water" cuts no epistemic ice.

I've said that getting a lot of information of the form the so and so is *F* is the warrant for producing sentences of the form "*A* is *F*" and that this supports the view that how things are being represented to be is that the

[7] In presenting this material in the past I have come across two misunderstandings. One is that the watery stuff (in David Chalmers's term) is watery everywhere it appears. There is a bit to say about how "the watery stuff" might be cashed out (see Jackson 2003) but on any viable cashing out, what is required is that the stuff be watery in some places where it appears—enough to make good sense of the way the term "water" got into the language; but this allows for the obvious fact that water is often not watery. Second, there is no disagreement here with the importance of causation: we interact equally with, and our brains equally carry traces of, H_2O and the watery stuff.

D is *F*. This says nothing about how we might determine *D* for any given case. We have given an argument for the representational content being of a certain shape without giving specifics. Let me now describe a way we might tackle this issue.

VII.

Take the sentence "Shakespeare wrote *King Lear*" (let's read the predicate "adjectively" to save complications). The credence researchers give this sentence has waxed and waned over the years. (It is very high at the moment.) Imagine a researcher borrowing the *Tardis* and travelling back in time to the early seventeenth Century. Her observations will be of the following kind: a certain human body stands in so and so a relation to an emerging text that opens "I thought the king had more affected the Duke of Albany than Cornwall"; the body interacts with other bodies in ways that produce words like, though not quite the same as, "Shakespeare" and, maybe, "Marlowe" or "Jonson"; the body stands in certain relations to the putting on of plays; it and the bodies and writings and plays around it are at one end of long information-preserving causal chains that end in current texts that contain words like "Shakespeare", "King Lear" and "Marlowe"; and so on and so forth. Like all of us (to repeat a point made earlier) our researcher will have no "thisness" detector. Her grounds for holding that she is, or is not, observing Shakespeare will be one and all grounded in observations of the distribution of properties and their causal interconnections. Her credence in "Shakespeare wrote *King Lear*" will, in consequence, be a function of the many possible observations of this kind. Descriptivists can then identify what she affirms in the sense of how she represents things to be using the sentence "Shakespeare wrote *King Lear*", as that the so and so wrote *King Lear*, for the value of "so and so" such that the credence of the so and so's writing *King Lear* takes those values at those observations, and "so and so" will be the associated description or property for her for the name "Shakespeare".

In general we can say that "*A* is *F*" in *S*'s mouth represents that the *D* is *F* when the credence profile under the impact of information of "*A* is *F*"'s being true matches that of the *D*'s being *F.*

VIII.

I now make a short digression to consider an objection sometimes put to me at about this point in the argument. It is granted that information to the effect that the so and so is *F* warrants us in making assertions of the form "*A* is *F*". But it is insisted that the warrant goes via a theory of reference that is *a posteriori*. We in effect have a two premise argument: facts of the form "The so and so is *F*" *plus* an *a posteriori* theory of reference *together* justify the assertion of "*A* is *F*".

But what might this theory of reference look like? If it says that sentences of the form "*A* is *F*" represent that the *D* is *F*, we are in agreement. There is only an objection if some other theory of reference is in the offing which implies that information of the form "The so and so is *F*" does not in itself warrant saying "*A* is *F*". But then it would make sense for philosophers of reference to warn those historians, journalists and politicians who go around producing sentences like "Mark Twain is Samuel Clemens", "Shakespeare wrote *King Lear*", and the like, on the basis of information of the form "the so and so is *F*", that because it is an *a posteriori* question what the right theory of reference is, they may have to retract their claims in the light of philosopher's investigations into the theory of reference. I think the journalists and historians would think it bizarre to be receiving instructions from philosophers *qua philosophers* on how likely it is that the sentence "Mark Twain is Samuel Clements" is true, or to be told that there is some *extra* that they need to know about over above the distribution of properties to settle the truth of "Mark Twain is Samuel Clemens".

IX.

I have been arguing that names have associated properties. I have said very little about what those properties might be for any given name. From the perspective of the description theory, there cannot be any generally applicable answer to this question. As Kripke says in a passage I often quote:

> The picture that leads to the cluster-of-descriptions theory is something like this: … one determines the reference for himself by saying—'By Gödel I shall mean the man, whoever he is, who proved the incompleteness of arithmetic'. Now you can do this if you want to. There's nothing really preventing it. You can just stick to that determination. If that's what you do, then if Schmidt

> discovered the incompleteness of arithmetic you *do* refer to him when you say, ‘Gödel did such and such’. (1980: 91)

Kripke is reminding us that we are free to use names as short for definite descriptions and, in particular, are free to use “Gödel” for the person who proved the theorem. This is not how we use “Gödel” but it might have been. Warped across to the description theory of names, the message is that there are many possible choices to be the property associated with any given name. However it makes good sense that often we chose certain causal-informational properties. The reason is the role of many sentences containing names in giving information. I will close with some comments on this important role for such sentences. There is a much longer discussion in Jackson 2005.

Token sentences in newspapers, books and computer screens and so on, like “Hackett won last night”, “Blair met Bush yesterday”, “Shakespeare wrote *King Lear*” and “There are floods in California” are sources of information (or sometimes misinformation), and their function as sources of information depends, as the folk know, on the existence of a causal information-preserving chain from, as it might be, Hackett’s winning and the appearance of the sentence token in the newspaper or the wave pattern token issuing from the radio. It is very useful to have such sources of information[8] and, in consequence, we should expect the most plausible candidate for the associated description or property for a name “*A*” to be, in many cases, something like *the thing of such and such a kind at the far end of an information-preserving causal chain ending in a certain token of* “A” *in a certain sentence token of, say, the form* “A *is* F”. This is why we should expect a degree of agreement between description theorists and causal theorists about the importance of causation to reference, though there remains the important difference over whether or not the causal descriptions and properties are *associated* with the names.[9]

[8] For more on the informational value of names, see Jackson 2005, but the point goes back a long way, see, e. g., Searle 1983: Ch. 9.

[9] I am indebted to the discussion at ANU in August 2005 and at Erfurt in September 2005.

REFERENCES

Bennett, J. (1976). *Linguistic Behaviour*, Cambridge University Press, Cambridge.

Grice, H. P. (1957). 'Meaning', *Philosophical Review* 66, 377–388.

Jackson, F. (1998). 'Reference and Description Revisited', *Philosophical Perspectives* 12, *Language, Mind, and Ontology*, (ed.) James E. Tomberlin, Blackwell, Cambridge, Massachusetts, 201–218.

— (2003). 'Narrow Content and Representationalism—or Twin Earth Revisited', 2003 Patrick Romanell Lecture, *Proceedings American Philosophical Association* 77, 55–71.

— (2005). 'What are Proper Names For?', *Experience and Analysis*, Proc. 27th International Wittgenstein Symposium 2004, (ed.) Johann C. Marek and Maria E. Reicher, hpt-öbv, Vienna, 257–269.

Kripke, S. (1980). *Naming and Necessity*, Basil Blackwell, Oxford.

Lewis, D. (1969). *Convention*, Harvard University Press, Cambridge, Massachusetts.

Locke, J. (1690). *An Essay Concerning Human Understanding*, Book III, Ch. II, § 2.

Putnam, H. (1975). 'The Meaning of "Meaning"', *Mind, Language and Reality,* Cambridge University Press, Cambridge.

Searle, J. (1983). *Intentionality*, Cambridge University Press, Cambridge.

Soames, S. (2005). *Reference and Description*, Princeton University Press, Princeton.

Grazer Philosophische Studien
74 (2007), 251–272.

CONTEXTUALISM ABOUT KNOWLEDGE AND JUSTIFICATION BY DEFAULT

Marcus WILLASCHEK
Universität Frankfurt/M.

Summary

This paper develops a non-relativist version of contextualism about knowledge. It is argued that a plausible contextualism must take into account three features of our practice of attributing knowledge: (1) knowledge-attributions follow a default-and-challenge pattern; (2) there are preconditions for a belief's enjoying the status of being justified by default (e.g. being orthodox); and (3) for an error-possibility to be a serious challenge, there has to be positive evidence that the possibility might be realized in the given situation. It is argued that standard "semantic" versions of contextualism (e.g. those of Lewis, Cohen, DeRose) fail to take these features into account, which makes them overly hospitable to the sceptic, and that Williams' version of contextualism, although incorporating (1), fails to do justice to (2) and (3). According to the contextualism developed here, although epistemic standards vary with the context, the truth-value of particular knowledge-attributions does not. Contexts here are understood as being constituted by two elements: an epistemic practice (a rule-governed social practice such as a scientific discipline, the law, a craft etc., in which knowledge-claims are evaluated according to specific standards) and the "facts of the matter" (i.e. those facts which, together with the epistemic standards in question, determine which error-possibilities are relevant and thus have to be eliminated for a knowledge-claim to be true). If there are several epistemic practices, and thus several contexts, in which a knowledge-claim can be evaluated, it is the "strictest" practice that counts. In this way, the counterintuitive consequence of other versions of contextualism that the same knowledge-claim can be true in one context, but false in another, can be avoided. At the same time, scepticism can be resisted since even in the "strictest" epistemic practices, error-possibilities become relevant only when backed by positive evidence that they might in fact obtain.

Contextualism about knowledge is the view that the standards someone must meet in order to know something vary with the context of ascription. In this paper, I want to defend and refine a contextualist approach to knowledge and scepticism. After a brief exposition of the sceptical problem, I will sketch the standard contextualist approach to it as expressed (with significant variations), for instance, in the work of David Lewis, Steward Cohen, and Keith DeRose, and argue that this approach is unconvincing, among other reasons because it is too hospitable to the sceptic (1). Looking at knowledge-attributions in real-life cases will motivate a contextualist approach enriched by a "default and challenge" conception of justification (2), as has been proposed before by, among others, Michael Williams. Although I sympathise with much of Williams' account, I will argue that his conception of a "default justificational status" is insufficiently complex, and that, for this reason, his version of contextualism is also overly hospitable to the sceptic (3). Next, I will sketch some features of a sufficiently complex contextualist-cum-default and challenge conception of knowledge. Perhaps the most remarkable feature of the contextualism defended here, as compared to other versions of contextualism, is that it does *not* imply that the same knowledge-claim can be true in one context but false in another. This in turn is a consequence of construing contextualism about knowledge not as a linguistic thesis about the usage of the expression "to know" and its cognates, but rather as a claim about the different standards at work in different epistemic practices. As I will argue, if a knowledge-claim can be evaluated by the standards of different practices, it is always the "strictest" practice that counts (4). Despite incorporating some features of "absolutist" conceptions of knowledge, however, the "epistemic practice" contextualism defended here can deal with the sceptical challenge in a satisfactory way (5).

I.

I leave my apartment. In the staircase, I stop and ask myself whether I locked the door. Did I? Do I know that I did? In order to know that I did, I must be able to rule out that I forgot to lock the door, which on reflection I can't. So I don't know that I locked the door.—Now I turn back to check whether I locked the door. I press the handle and find the door is locked. Do I *now* know that the door is locked? The intuitive answer clearly is: yes, now I know the door is locked.

But many epistemologists would hesitate. After all, there are many possible situations compatible with my finding the door locked (more precisely, with my believing the door is locked after pressing the handle) in which the door is not locked. Here are some such situations:

(a) The handle might be stuck.
(b) I might not have pressed hard enough.
(c) Someone might have tricked me by holding the handle on the other side of the door.
(d) I might only have dreamt to have checked the handle.
(e) I might only have dreamt to have checked the handle because everything I seem to experience is really part of a dream.
(f) There might be no door, no handle, no hands, since I am a deluded brain in a vat.[1]

And, one could argue, as long as I can't rule out (a)–(f), I don't know that the door is locked.

Now possibilities (a)–(f) clearly fall into two distinct categories: while (a)–(d) can be empirically checked, and thus it is possible to rule them out at least in principle, (e) and (f) can never be ruled out by empirical investigation. (e) and (f) are so-called "sceptical possibilities": If they obtain, there is no evidence that they do obtain; if they don't obtain, there is no evidence that they don't obtain, because all possible evidence is equally compatible with both their obtaining and their not obtaining.

Many philosophers think that sceptical arguments of the following type have a strong intuitive plausibility (the "argument from ignorance")[2]:

(1) I don't know that I am not a brain in a vat.
(2) If I don't know that I am not a brain a vat, then I don't know that I locked the door.
(3) I don't know that I locked the door.

Since the same argument can be run for any proposition about the so-called external world, it seems to follow that we really don't know most of what we take ourselves to know.

One account of why this argument seems so compelling is offered by the sceptic: the argument seems compelling because it is correct. It is deductively valid, and its premises and conclusion are true. We really

[1] Michael Williams offers a structurally similar list (concerning travel guides) in Williams 2003.

[2] DeRose 1995.

cannot know such things as whether the door is locked, whether I have two hands, and so on. However, almost everyone thinks that scepticism is false. So we have to explain not only why the sceptical argument seems compelling, but also where it goes wrong. Most of the various kinds of contextualism that have been proposed over the last 25 years aim primarily at meeting this double explanatory demand.

One prominent version of contextualism has been developed in different ways by David Lewis, Keith DeRose, and Steward Cohen.[3] According to this view, which has been called "semantic" contextualism,[4] the context to be considered in evaluating knowledge-claims and knowledge-attributions is the conversation in which the claim is put forward or the attribution is made. On this view, there are two basic kinds of contexts: ordinary and sceptical. Ordinary contexts are characterized by low epistemic standards, which means that in order for *S* to know that *p* only some error-possibilities must be ruled out, namely those which are considered to be relevant in the context of the particular conversation in which knowledge is ascribed.[5] (An error-possibility for the belief that *p* is any possibility in which non-*p*.) In a sceptical context, by contrast, epistemic standards are high: in order for *S* to know that *p all* alternatives to *p* must be ruled out. Most conversations generate an ordinary context, but once sceptical scenarios like the brain in a vat scenario are posited, a sceptical context is created; the epistemic standards change from low to high, with the result that all ordinary knowledge-claims turn out to be false in the context of that conversation. Note that this means that while two subjects S_1 and S_2 may possess the same evidence with respect to their beliefs that *p*, S_1 may know, and S_2 may not know, that *p*, if the respective contexts of attribution differ. Even more paradoxically, this means that a statement of the form '*S* knows that *p* at time *t*' may be true in one conversational context, but false in another, even if everything about the subject, the available evidence, and the facts of the matter are held constant.

In ordinary contexts, in which sceptical hypotheses are not considered, we may easily know such things as that we locked the door. All we have to do is, for instance, remember that we locked the door, and check or double check if we feel uncertain. But once sceptical hypotheses are considered, we enter a different context in which remembering or checking

[3] Cf. e.g. Cohen 1988, DeRose 1995, Lewis 1996.

[4] Cf. Prichard 2002.

[5] This formulation is meant to be neutral with respect to internalist (Cohen) and externalist (Lewis, DeRose) versions of semantic contextualism.

is not good enough. In effect, in this sceptical context nothing is good enough. Therefore, in a sceptical context, we do not and cannot know that we locked the door (or know anything else about so-called external reality).

The intuitively compelling character of the sceptical argument is thus explained in the very same way as it is explained by the sceptic: the argument is compelling because it is valid and its conclusion is correct. On the other hand, however, full-blown scepticism is avoided because the victory of the sceptic is limited to sceptical contexts: in ordinary contexts, where sceptical hypotheses are not at issue, we continue to know what we ordinarily take ourselves to know, including the fact that we locked the door. But even though scepticism is avoided, according to semantic contextualism the sceptic wins every argument simply by mentioning sceptical scenarios. I now want to argue that this can't be right.

II.

Let us return to the list of possibilities supposedly incompatible with my knowing that I locked the door. As mentioned before, these possibilities fall into two classes: those that can at least in principle be empirically ruled out and those that can never be ruled out. The latter, of course, are the ones that give rise to sceptical problems. Now imagine again that I have just left my apartment, but that this time I am not alone, but with a friend. Halfway down the stairs the friend asks me whether I locked the door. Let us consider three different cases:

> *Case a*: I answer "Yes", and that is the end of it. Since an unqualified assertion typically expresses a claim to knowledge,[6] by answering "Yes" I claim to know that I locked the door. If this claim goes unchallenged, no further argument or reason-giving is needed. (Whether this means that no evidence is needed at all is a question to which I will return.)
>
> *Case b*: Again, my friend asks whether I locked the door, but this time I am not sure and return to check. Again I find the door is locked. But my friend is not satisfied and suggests: "The handle might be stuck".

[6] Williamson 2000: 11f.; cf. Williams 2001, 27.

> How would I react? It depends. If the handle had been stuck before, I might say: “Yes, you’re right, I’d better check the handle”. In that case, I would perhaps unlock the door with my key, try the handle when the door is open, etc. But what if the handle never got stuck before, or if I fixed it only recently? Then the rational answer to the suggestions that the handle might be stuck would clearly be: “No, don’t worry, the handle’s okay. The door is closed all right”.
>
> *Case c*: My friend asks whether I locked the door, I say yes, she asks how I know, I respond that I clearly remember that I locked the door only a minute ago, and now my friend objects: “But perhaps you only dreamt that you locked the door”. What would I say to this? Probably my first reaction would be one of puzzlement: “What do you mean? How can you think that I only dreamt I locked the door when you were with me when we left the apartment?” Now again there are two possibilities: Either my friend has a good answer to this question; for instance, she might point out that I am severely sleep-deprived and apt to fall into second-long periods of sleep without my noticing. In this case I might return to check the door, thereby admitting that I do not really know that the door is closed. But what if my friend has no such answer to give? What if the only answer comes to this: “It is logically possible and compatible with all available evidence, including our apparent memories to the contrary, that you only dreamt that you locked the door”. Would I, should I, return and check the door? The obvious answer is no. I would neither return, nor would it be rational for me to do so. Rather, I would insist that I did lock the door, thereby claiming to know that I did. And the same obviously goes for the two sceptical scenarios—that I might be dreaming all the time and that we all might be deluded brains in vats: if this is all my friend had to offer in order to question my claim to knowing that I locked the door, the only rational reaction would be to point out that this may be logically possible, but that these far-fetched possibilities are irrelevant unless backed by further considerations. In short, the rational reaction is to insist without any further argument that I know that I closed the door.

So we find that, at least in ordinary, non-philosophical contexts, there is a distinction between error-possibilities we do (and ought to) take seriously and others we don’t (and needn’t). What is missing in the cases in which we don’t take an alternative seriously is what Peirce once called a “positive reason"[7] for doubting the belief in question. If I am sleep-deprived, or if, as in the earlier situation, the handle is old and tends to get stuck, these are positive reasons for doubting. Put differently, in order to question my belief that I locked the door, what is needed is *evidence to the contrary*. But a mere logical possibility is no evidence at all, neither for nor against something. It is logically possible that unbeknownst to me I just inherited a fortune from an uncle whom I have never heard of before. But the fact that this is logically possible is no evidence at all for the belief that I just inherited a fortune. To be sure, the logical possibility of *p* can rule out one kind of objection against the belief that *p*, namely that *p* is logically impossible. But this can hardly count as evidence in favour of the belief that *p*. Hence, the mere fact that it is logically possible that *p* can never serve as evidence for the belief that *p*—and, as the examples just considered show, it can also not serve as evidence against the belief that not-*p*.

There are three lessons I want to draw from the examples just considered.

First, our real-life practice of putting forward, questioning, and justifying knowledge claims follows what Robert Brandom and others have called a *default and challenge* pattern[8]: in many situations, knowledge-claims are considered to be legitimate without any backing by explicit justification or reason-giving. However, this default status is open to challenges.

Second, for a knowledge-claim to enjoy default status, it must meet certain conditions. Here contextual factors come into play. For instance, knowledge-claims that agree with general and/or expert opinion will typically enjoy default status, while heretical claims require explicit justification. Also, a claim about whether or not *p* issued in situations that are conducive to correct judgement about *p* will typically be default, while claims about something that is out of the subject's standard range of epistemic access will need explicit backing. If I claim that I just locked the door to my apartment, then this claim will typically enjoy default status, while my claim that Angela Merkel just locked the door to her office in Berlin will not. The reason is that memory affords a standard means of epistemic access to what *I* just did, but not to what someone else just did far away from me.

What is remarkable about this is that the deliverances of memory do not play the role of evidence: whether I really remember that I locked the

[7] Peirce 1868: 140.

[8] Brandom 1994: 177; Williams 2001: 25.

door is irrelevant for my claim that I did lock the door to enjoy default status. What counts is that generally people *do* remember correctly—and thus may be *supposed* to remember correctly—what they did a minute ago. Whether I really remember or not becomes relevant only when the reliability of my memory is being questioned. (The same holds for other modes of knowledge-acquisition such as perception and testimony.) I will return to the question of preconditions for enjoying default status in more detail below.

The third and final lesson I want to draw is this: not any old logical possibility can serve as a challenge to a knowledge-claim that enjoys default status. *It is simply not true that the mere mentioning of error-possibilities undermines a claim to knowledge.* This means that the semantic contextualist diagnosis of scepticism can't be correct, because it assumes that mentioning or thinking of sceptical scenarios suffices to raise the epistemic standards in such a way that we no longer know what we do know in ordinary contexts. What is needed is more than that, namely evidence to the contrary: either evidence that the claim is in fact false or evidence that the conditions required for default status are not fulfilled. For instance, the fact that there is no non-circular argument for the reliability of memory as a source of information about the past is not enough to challenge my claim that I just locked the door. Rather, what is needed is evidence that *my* memory, in this very situation, is not reliable. Again, contextual factors come into play, since the distinction between relevant and irrelevant challenges, challenges that must be answered and those that can reasonably be shrugged off, depends on the given situation and specific circumstances: For someone suffering from a loss of short-term memory, the question of whether one really remembers to have locked the door poses a serious challenge; to someone whose memory works properly, it does not.

Notice that this is not just externalist reliabilism: What depends on the context (e.g. on the reliability of someone's memory) is not simply the question of whether the person knows that *p* or not, but rather what standards the person must meet in order to know that *p*: is it enough for the person in the given situation to correctly believe that *p* (knowledge by default), or must she adduce further evidence and rule out contrary evidence?

III.

A version of contextualism that has taken to heart the first lesson (about the default and challenge structure) is the one developed by Michael Williams, most clearly in his 2001 book *Problems of Knowledge*.[9] Williams keeps to the traditional conception of knowledge as justified true belief, but interprets it within a default and challenge model of justification, in which both default status and the relevance of challenges depend on various contextual factors.

Williams distinguishes between two kinds of justification: personal and evidential.[10] Personal justification concerns the question of whether the epistemic subject acts epistemically responsibly in believing and claiming what she does. Evidential justification, by contrast, consists in the "adequate grounding" of a belief by the available evidence. Both personal justification and objective well-groundedness are required for a belief to count as knowledge. The default and challenge model primarily concerns the subject's personal justification: I am personally justified in believing that *p* either if my belief that *p* enjoys default status or if I can counter all relevant challenges (including standing objections to *p*). As standard externalist scenarios such as "barn façade county"[11] show, personal justification of a true belief is not enough for knowledge: if I correctly believe that there is a barn over there, my belief may count as being justified by default, but if what I see is the only real barn among many façades, I still do not know that I see a barn. So Williams additionally requires what he calls the "adequate grounding" of a belief, which depends in part on the external circumstances in which the belief is formed and held (Williams 2001: 162).

Williams' version of contextualism differs from semantic contextualism in yet another important respect. Whereas the latter view traces all changes in epistemic standards to moves in a conversation, Williams takes into account a broad variety of contextual factors.[12] Since all of these factors can

[9] Williams 2001.

[10] Williams 2001: 22.

[11] Cf. Goldman 1976.

[12] Williams brings them under five types (cf. Williams 2001, ch. 14): (i) semantic (What must be presupposed for the question of whether *p* or not *p* to make sense/to arise?); (ii) methodological (What must be presupposed for our methods of inquiry to work; e.g. the reality of the past with respect to methods of historical research?); (iii) dialectical (What challenges are actually being advanced? Which are generally considered to be standing objections?); (iv) economic (How important is the correctness of the knowledge-claim to the subject and/or the attributor?

influence the epistemic standards according to which a given knowledge-claim must be evaluated, there is no clear-cut distinction between high and low standards, but rather a complex net of standards, any of which can be strict in one respect but rather loose in others.

Obviously, the default and challenge model of justification plays a central role in Williams' complex version of contextualism. But what about the other two lessons that can be drawn from the examples I considered?

The second lesson was that in order to enjoy default status a belief has to fulfil certain conditions, such as being orthodox, being formed under standard conditions for the reliable working of the chosen method of belief formation, and so on. By contrast, Williams holds that "one is entitled to a belief or assertion [...] in the absence of appropriate 'defeaters'"[13]. So a belief is justified by default, according to Williams, as long as it is not faced with serious challenges. Of course, one might argue that the heterodox character of a belief, or the fact that it was formed under non-standard conditions, should count as an "appropriate defeater". But this seems to me to mislocate their epistemological impact: The fact that in Copernicus' time almost everyone believed that the sun moves around the earth is no "appropriate defeater" to Copernicus' claim that the earth moves around the sun; if it were, Copernicus would have had to convert the majority of his contemporaries to a heliocentric view in order to know that the earth moves around the sun. The heterodox character of Copernicus' claim simply means that his claim did not enjoy default status and that, in order for his claim to count as knowledge, Copernicus had to *argue* for his view. Among other things, he had to give appropriate answers to the substantial objections raised by his contemporaries. One such objection, for instance, was that one ought to expect the movement of the earth to produce an airstream; but there is no such airstream (an objection answered convincingly only by Galilei). The fact that most people *believed* that the sun moves around the earth, by contrast, was no such substantial objection.[14] What this shows is that being heterodox rules out default status not in virtue of being a challenge or defeater, but rather because orthodoxy of a belief is a

How much effort is necessary to rule out certain error-possibilities?); and (v) situational (Is the chosen method of belief-formation in fact reliable in the given situation?).

[13] Williams 2001: 149.

[14] Obviously, there is no clear-cut distinction between objections that are and those that are not substantial. In general, substantial objections tend to focus on the specific content of the claim in question, while considerations that undermine a claim's default status tend to concern the subject's epistemic and dialectical position. But there may be exceptions.

prerequisite of its enjoying default status in the first place.

The same seems be true about the proper working of the method of belief-formation one employs. Imagine I look out of the window in plain daylight and correctly claim that it's snowing. Someone watches me and asks, "How do you know?" Assuming the person saw that I looked out of the window, I wouldn't know how to respond except by saying something the person already knows, namely that I know it is snowing because I looked out of the window. And it seems that in the absence of contrary evidence I am not epistemically required to give any further explanation of how I know that it is snowing.—But now imagine that I sit in a room with the shutters closed. Last time we looked outside it wasn't snowing. All of a sudden I claim that it's snowing. You ask: "How do you know?" or "How can you tell?" In this case, I am obviously required to give an informative explanation in order to be counted as knowing that it is snowing. It would not suffice to appeal to something the other person already knows, such as: "Well, I have been sitting in the room all the time; so of course I know that it just started snowing". What is called for is further information about my method of belief-formation. For instance, I might respond that I have a scar that itches every time it starts snowing. If I can't produce a convincing answer, I must withdraw my claim to know that it is snowing. Since the same bare question is being asked in both cases ("How do you know?"), and since in both cases no evidence against my claim is put forward, it is hard to see why the question should be a serious *challenge* in the second case but not in the first. The difference rather seems to be that in the second case, but not in the first, the question is meant to remind me that in the given situation my claim does not enjoy default status, but must be backed by explicit reason-giving.

Hence we must distinguish between two ways in which a belief can fail to enjoy default status: (1) by not meeting the requisite conditions such as being orthodox or being formed under standard conditions, and (2) by being faced with serious context-specific challenges.

I now turn to the third lesson I mentioned, namely that challenges to claims that enjoy default status need to be backed by evidence against the challenged claim, and that mere logical possibilities do not constitute evidence. Williams is aware of this: "A defeater does not come into play simply in virtue of being mentioned: there has to be some reason to think that it might obtain".[15] On the other hand, however, he claims in the same book

[15] Williams 2001: 161; cf. 150f.

that challenges based on sceptical hypotheses, once put forward, “deprive ordinary knowledge-claims of their default justificational status”,[16] and even endorses the view that “we may temporarily lose our knowledge when we project ourselves into the rarified context of ‘doing epistemology’”.[17] But a sceptical hypothesis is such that there *can* be no evidence either for or against it, since its obtaining and its not obtaining are equally compatible with all possible evidence. So with respect to a sceptical hypothesis, there is never a “reason to think that it might obtain”. It therefore seems to me that Williams is not consistent in claiming on the one hand that challenges must be based on evidence (“reasons to think that they obtain”) and conceding on the other that sceptical hypotheses “deprive ordinary knowledge-claims of their default justificational status”. We should stick with the former claim and reject the latter.[18]

IV.

Let us pause for some reflections about the general characteristics of the contextualism that emerges out of the critical discussion of semantic contextualism on the one hand and Williams’ version of contextualism on the other. First, the resulting kind of contextualism is not primarily a linguistic thesis about the correct use of the expression “to know” and its cognates. What depends on the context is the correctness of attributions of knowledge and justification, but these attributions typically take a non-linguistic form. If I hail a taxi, for instance, I attribute to the driver the knowledge that my behaviour means that I need a taxi. One might object that all I need to attribute to the driver is the corresponding belief. But if the default and challenge account sketched above is correct, then the driver’s belief will typically be justified by default; and since the belief in question is true, there is no reason not to regard it as knowledge. If I were to give linguistic expression to what I attribute to the driver, it would be

[16] Williams 2001: 186. Patrick Leland has suggested to me that this statement might be read more charitably as expressing not Williams’ own view but the view of the Cartesian sceptic. But even if that is granted, there remains the fact that Williams explicitly endorses the claim that knowledge is instable (cf. fn. 17).

[17] Williams 2001: 195; also cf. Williams’ discussion of the “instability of knowledge” in Williams 1991. In a similar way, David Lewis has claimed that knowledge is “elusive”: We have it only as long as we don’t reflect on it.

[18] But cf. Williams 2003: 990f., where Williams himself criticises contextualists such as Lewis and DeRose for being too hospitable to scepticism.

most natural to say that, being a taxi driver, he *knows* what my hailing a taxi means. So much for non-linguistic attributions of knowledge. But even attributions that take a linguistic form will typically not employ the expression "to know" and its cognates.[19] If I ask you what time it is and you answer "It's two o'clock", you implicitly claim to *know* that it's two o'clock. Otherwise, you ought to have said something like: "I'm not sure, but I believe it's two o'clock". As mentioned before, straightforward assertions that *p* typically have the force of self-attributions of knowledge that *p*. And if I react to your assertion by going to the train station in order to catch a train at half past two, then I accept your knowledge-claim and thus attribute knowledge to you. What depends on the context is the correctness of knowledge-attributions in this sense (and thus the conditions for knowledge). Only a small fraction of knowledge-attributions take the form of sentences containing the expression "to know". Thus, the contextualism I want to defend is not primarily a linguistic thesis. As I want to show now, it also does not imply that the truth-value of knowledge-attributions, linguistic or not, *varies* with the context.

In order to see why this is not the case, it is necessary to say a bit more about what a context is. We have already rejected the idea that the kind of context relevant for knowledge-attributions is simply the context of a conversation. I now want to suggest that the context on which the correctness of knowledge-claims depends is constituted by two factors: an epistemic practice and the relevant facts of the matter. Epistemic practices are rule-governed social practices in which knowledge is acquired and attributed according to specific epistemic standards: sciences such as biology or physics, practices such as law, medicine, engineering, etc. Besides these highly regimented and self-reflective epistemic practices there is the vast variety of social practices such as crafts, commerce, and sports, each of which has some specific epistemic standards of its own, but which mostly employ the same all-purpose set of epistemic standards that governs commonsense attributions of knowledge. Let's call this latter our "ordinary" epistemic practice. The epistemic standards employed in different practices overlap, but there are also important differences. In the empirical sciences, for instance, knowledge is tied to the possibility of empirical confirmation; in mathematics and related formal disciplines, knowledge requires proof; in the law, knowledge from testimony is restricted by certain formal procedures such as taking an oath; in various crafts, practitioners can tell

[19] Cf. Williams 2001, 27.

things apart simply by looking or touching, while laypersons can do so only by indirect methods; etc. etc. These standards determine (a) what kinds of beliefs enjoy default status; (b) what kinds of error-possibilities are relevant challenges; (c) what counts as answering a challenge (as ruling out an error-possibility); and (d) what counts as establishing a claim that does not enjoy default status. However, these standards are *general* in the sense that they must be applicable to a broad variety of cases. In order to determine the epistemic status of a *specific* belief in a given situation, something else must be taken into account—namely, the relevant facts of the matter.

By this I mean those facts on which it depends *which* error-possibilities satisfy the criteria of relevance of a given epistemic practice in the given situation. For instance, before there were fake Rolex watches on the market, one could know that the watch before one was a Rolex watch simply by reading the brand name on the watch face; the mere logical possibility that someone might counterfeit Rolex watches was no relevant challenge as long as there was no indication that someone really did so. However, if there are thousands of fake Rolex watches on the market, the possibility that the watch before me is a fake becomes relevant—even if I don't know about the existence of fake watches. If I cannot tell the fake from the real thing and if there is a significant number of fakes around, then I simply don't know that this watch before me is a Rolex (assuming that all I have to rely on is my own judgement). I suggest that we analyse this situation as follows: there is a general standard implicit in our ordinary epistemic practice according to which one knows that *p* only if one can rule out all *relevant* error-possibilities. Obviously, the specific criteria of relevance will vary with the epistemic practice and the facts of the matter. Here are some examples of criteria of relevance implicit in our ordinary epistemic practice: A possibility's being relevant requires *some* reason to think that it in fact might hold; merely logical possibilities are never relevant in this sense; for an error-possibility to be relevant, its obtaining must be consistent with generally accepted knowledge as well as with specific knowledge about the subject's situation; an error-possibility is relevant if the kind of error in question is common or to be expected under the circumstances; etc. etc. Now the facts of the matter determine, with respect to the case at hand, *which* error-possibilities satisfy the criteria of relevance specified by the epistemic standards: in the situation we imagined, the possibility that the watch is a fake is relevant since there are many fakes out there and thus some reason to think that the watch might be a fake. By contrast, the possibility that one is only hallucinating a watch is not relevant as long as there is no reason to think that one is hallucinating. In the imagined situation, then, one knows that one holds a watch in one's hand, but one does not know that it's a Rolex. The epistemic standard (rule out all error-possibilities for which there is reason to think that they may in fact obtain) and the facts of the matter (many fakes out there) taken together thus determine whether an error-possibility is relevant or not. In this sense, knowledge *depends* on context.

I will call this kind of contextualism "epistemic practice contextualism", because it is epistemic practices that set the standards for knowledge and justification. However, this is not to deny that a specific context also includes the facts of the matter that determine which error-possibilities are relevant in the given case. So changes in the relevant facts constitute changes in context if the factual changes affect the relevance of error-possibilities—even if the epistemic practice remains the same.

That knowledge depends on context, however, does not necessarily mean that the truth of individual knowledge-claims *varies* with the context, so that the same claim can be true in one context and false in another. The reason for this is that the correctness of a knowledge-attribution in *any* context typically depends on the standards of the *strictest* context in which the claim can be evaluated. Let's call this latter context "the standard-setting context". In the case of Rolex watches, for instance, the standards are set by the context of watch-making: if an expert watchmaker opens the watch, dissembles the clockwork and finds that the watch is a genuine Rolex product, then the possibility that the watch is a fake has been effectively eliminated. No further confirmation is needed, and none is possible. We cannot increase our confidence, for instance, by handing the watch over to a physicist and asking him to check if it's genuine. Even though the epistemic standards employed by the physicist may be in some respects "stricter" than those of the watchmaker, this typically won't help in determining whether the watch is genuine or not, since physicists, as such, are not experts in distinguishing genuine from counterfeit watches.[20] This suggests that the epistemic standards with respect to a given knowledge-claim about some subject matter *M* are set by the "expert practice", if there is any—that is, by the practice of those people who know best about *M* and to whom one defers judgement in that matter.

[20] Of course, there may be exceptions—perhaps a fake can be distinguished from the real thing only on a molecular level. What this shows is that the assignment of knowledge-claims to epistemic practices, too, depends on the relevant empirical facts of the matter.

Whether someone knows that *p* thus depends on whether the assertion that *p* is justified according to the epistemic standards of the "expert practice". Otherwise, the layperson, simply because her standards are less exacting, could know things about a scientific subject matter that experts do not know, which seems absurd. (Of course, the difference between experts and laypersons does not only concern epistemic standards, but also, and even more importantly, the ability to determine the facts of the matter.)

However, a layperson need not become an expert in order to know what the experts know, since implicit in our ordinary epistemic practice is the standard that one is justified in one's beliefs about a subject matter *M* if one bases one's beliefs about *M* on what is considered as knowledge about *M* by the respective experts. The expert justification of the belief need not be available to the layperson; still, her being justified in her belief about genuine Rolex watches, supernovae, or the human genome ultimately depends on the belief's meeting the epistemic standards of the expert practice. Such are the benefits of the epistemic division of labour.

Note, by the way, that sceptics, by definition, are not experts in any field. By admitting that the standards are set by the "strictest" practice in which a knowledge-claim can be evaluated, the epistemic practice contextualist does not surrender to the sceptic, since even the strictest epistemic practices distinguish between relevant and irrelevant error-possibilities and allow for default justification. As we've already seen, even the watchmaker typically need not be able to rule out that he is dreaming in order to know that a watch is a genuine Rolex. I will return to the issue of scepticism in the last section.

For many subject matters there are no experts—or, equivalently, everyone's an expert: It doesn't take a meteorologist to determine whether it's raining here and now; all it takes is average perceptual capacities. Here, the standard-setting context is just our ordinary commonsense practice of evaluating knowledge-claims. But even in this case, changing to the scientific context of meteorology does not undermine the validity of our ordinary epistemic standards for the question of whether one knows that it's raining here and now: if someone knows that it's raining here and now according to ordinary standards, it is hard to see how a meteorologist might prove her wrong (according to meteorological standards). The reason is that with respect to the question of whether it's raining here and now, meteorologists rely on the same epistemic standards as ordinary folk.

So the situation is this: either there is an expert practice with respect to subject matter *M*, in which case someone has knowledge about *M* only if her true belief is justified according to the standards of that practice. Or there is no expert practice, in which case someone has knowledge if her true belief is justified according to our ordinary epistemic standards. Either way, changing the context will not change the truth value of a knowledge-claim. Hence, the truth-value of knowledge-claims *depends* on the context, but doesn't *vary* with the context. What varies with the context are the epistemic standards according to which knowledge-claims are evaluated; but for every particular knowledge-claim there is precisely one context that sets the standards for that claim. So in one important respect epistemic practice contextualism is invariantist. Nevertheless, it is a kind of contextualism in the sense of the opening sentence of this paper, in that different epistemic practices constitute different contexts with different epistemic standards. Relativism, however, is avoided by recourse to the epistemic division of labour: If there are no experts, our ordinary standards count; if there are experts, however, their standards kick in. Obviously, there can be much controversy about who's an expert, what the relevant standards are, and whether or not a knowledge-claim meets the relevant standards. But these questions and the controversies about them make sense only if there can be correct and incorrect answers to them, which presupposes a non-relativist epistemology.[21]

But what then about an observation that seems to speak in favour of semantic versions of contextualism, namely that we tend to attribute knowledge rather generously in some contexts, but more parsimoniously in others? For instance, if Peter tells Mary today that he will go to Munich tomorrow, then it seems absolutely appropriate for Mary to say tomorrow that she knows that Peter is in Munich. (Imagine Peter and Mary work in the same office in Frankfurt and someone asks where Peter is and Mary answers: "I know where he is! He's in Munich. He told me himself". There seems to be nothing wrong with what Mary says.) But now imagine that later Peter is accused of a crime in Frankfurt that took place on the very day he was allegedly in Munich. It seems that Mary can no longer say that she knows where Peter was if all she has to rely on is his announcement that he was going to go to Munich. The reason is that now error-possibilities

[21] Therefore, the kind of contextualism I want to defend does not imply that explicit knowledge-ascriptions change their truth-value with changes in context. This means that the linguistic arguments against "semantic" contextualism (e.g. Schiffer 1996; Stanley 2004; Capellen/Lepore 2005) do not apply.

have become relevant that were not relevant before. So it seems that the correctness of a particular knowledge-claim, and not just the epistemic standards, can vary with the context after all.

In response, the epistemic practice contextualist can make use of what Keith DeRose has called a "Warranted Assertibility Maneuver"[22]: the weaker standards according to which Mary knows that Peter was in Munich do not concern the *truth-conditions* of her knowledge-claim, but only the conditions of *warranted assertibility* of knowledge. In other words, it was appropriate for Mary to *say* that she knew, even though in fact she didn't know. In many situations it may be appropriate for pragmatic reasons to ascribe knowledge without making recourse to the "strictest" standards. For instance, if nothing much hinges on whether Mary is right or not, or if checking is difficult, then her belief will pass as knowledge. But it seems that when we ask whether she "really" knows, pragmatic and economical considerations just don't count. Perhaps the possibility that Peter did not go to Munich originally *seemed* irrelevant to Mary, but then it turned out to be relevant after all. So what we should say is that, even though it became apparent only afterwards, Mary never knew that Peter was in Munich. (Again, this does not imply surrender to the sceptic as long as we hold fast to the idea that even the strictest standards for a given knowledge-claim require a distinction between relevant and irrelevant error-possibilities.)

Much more needs to be said in order to defend epistemic practice contextualism against its contextualist and non-contextualist rivals. Here my aim is only to give a rough sketch of a version of contextualism that promises to avoid the problems and weaknesses of both the standard semantic contextualisms and Williams' non-semantic version of contextualism. Working out the details of epistemic practice contextualism would primarily require a detailed account of the criteria of relevance employed in various epistemic practices, a project which would be of interest quite independently of most epistemologists' obsession with scepticism. Having said this, I will now return to the question of how to respond to the sceptical argument.

[22] DeRose 1995; of course, DeRose discusses "warranted assertibility maneuvers" as moves made by the critics of (semantic) contextualism. The epistemic practice contextualist can acknowledge the limited force of these since her view is not primarily about assertions, but about knowledge.

V.

What, if anything, is wrong with the "argument from ignorance"? Williams' diagnosis of scepticism is centred around the idea that the sceptic must presuppose what Williams calls "epistemological realism"[23]: the view that our beliefs and claims fall into natural epistemological kinds, like "beliefs based on perception", "beliefs based on memory", and so on, and that the beliefs stand in fixed evidential relations by virtue of the epistemic kind to which they belong. Only epistemological realism, so Williams claims, allows the sceptic to question all our knowledge-claims, or all our knowledge-claims of a certain kind, at once. Otherwise, the sceptic would have to proceed in a piecemeal fashion, investigating the epistemic standing of one individual claim after another, which of course means that the sceptic wouldn't get anywhere near the general conclusion that we don't know anything, or at least that we don't know anything in various broad domains.

This is an illuminating diagnosis, but it is difficult to see how it can be brought to bear on the sceptical argument considered above:

(1) I don't know that I am not a brain in a vat.
(2) If I don't know that I am not a brain a vat, then I don't know that I locked the door.
(3) I don't know that I locked the door.

Of course, this argument, if successful, only shows that I don't know that I locked the door. But no theory such as epistemological realism is needed in order to generalise this result; all we need is the class of propositions incompatible with my being a brain in a vat. Which propositions belong in that class is a matter of logic and linguistic meaning. And if we have a grasp on the class of propositions incompatible with my being a brain in a vat, we can generalise the argument simply by pointing out that, if successful at all, it works for all propositions of that class and shows that they cannot be the content of legitimate knowledge-claims.

So a different diagnosis is called for, and I think that the default and challenge model of justification affords such a diagnosis. To see this, consider a situation in which we put forward the first premise of the sceptical argument not as part of a syllogism, but as a challenge in a conversation:

[23] Williams 2001: 170f., 191ff.; cf. Williams 1991.

A taxi-driver claims to know how to get from here to Goethestrasse; you point out that this can't be right because he doesn't even know that he is not dreaming.[24] Assuming the taxi-driver is not, as it might happen, a former student of philosophy, and assuming further that he considers you worthy of an answer at all, he will not, and ought not, withdraw his knowledge-claim, but will rather insist that he knows the way alright and also knows that he's not dreaming. If you are a very insistent person, you may go on and explain to the taxi-driver why he can't possibly know that he's not dreaming: namely, because of the fact that any evidence for his being awake might in principle be a part of his dream. But if the taxi-driver has his wits about him, he will respond that he remembers how he woke up in the morning after a good night's sleep, how he went to work, and how you boarded his taxi and began a silly conversation. True, he might say, all this is compatible with the possibility of his being asleep and only dreaming this up, but this possibility in itself is no good reason to doubt his well-established belief that he is awake. The mere fact that it is conceivable that right now he is dreaming does not go against his being awake, as long as there is no positive evidence in favour of the possibility that he is dreaming.

Of course, it is highly improbable that you hit on a taxi driver who will give you a speech like that. But I think that this speech spells out what many people think when confronted with sceptical arguments: they simply don't see their point, and insist that they do know the sceptical hypotheses to be false. This is the first step of my diagnosis of the sceptical argument: I deny that it even *seems* to be compelling to non-philosophers. And the reason is not that non-philosophers "don't get it", but rather that they rightly and rationally insist on the principle that challenges to knowledge-claims—knowledge-claims that enjoy default status—require evidence against the correctness of the knowledge-claim. The mere possibility of an error, unaccompanied by any indication that it may actually have occurred, is no such evidence. So the first step of my diagnosis, if correct, achieves two things: It denies the first explanandum assumed by most philosophers, namely that the sceptical argument seems to be compelling; and at the same time explains the second explanandum, namely that the sceptical argument goes wrong somewhere. It goes wrong in the first premise, that we don't know the sceptical hypotheses not to obtain.

[24] I change the sceptical scenario from "brains in a vat" to "dreaming" merely to increase the plausibility of the narrative; challenging knowledge-claims in ordinary real-life situations by recourse to brain-in-a-vat scenarios would most probably evoke reactions of complete puzzlement—and questions about the mental sanity of the challenger.

Of course we cannot argue, in G. E. Moore-like fashion, from our having hands to our not being brains in a vat. But this is because having hands is no *independent* evidence for not being a brain in a vat, so that we can't argue from the one to the other.[25] But still we can know that we are not brains in a vat, because the belief that we are not enjoys default status, and no evidence against it has been offered. In fact, it is part of the idea of sceptical hypotheses that no evidence in their favour is even possible.

The second step of my diagnosis then must concern the question of why the sceptical argument seems so compelling to many *philosophers*; in particular, to explain why many philosophers find it plausible that we do not know the sceptical hypotheses not to obtain. The answer I want to propose is that in philosophy we abstract from the real-life contexts in which knowledge-claims are being issued and evaluated.[26] We don't ask ourselves whether this or that person in this or that situation knows the way to Goethestrasse, but whether anyone can ever know the way to Goethestrasse, or, for that matter, whether anyone can know not to be a brain in a vat. What we ask is whether, and how, anybody can know anything at all.[27] The effect of this de-contextualizing tendency of philosophical reflection is that there seems to be no distinction between the relevant and the irrelevant, between reasonable and unreasonable challenges, because these distinctions are highly context-sensitive. Once we abstract from context, all challenges can seem equally relevant. This is why I think it is important not just to consider sceptical syllogisms, but to imagine real-life situations in which knowledge-claims are being issued and criticised. This helps us keep in mind that there are error-possibilities which, because they are not backed by any evidence, are simply irrelevant for the question of whether we know the way to Goethestrasse. If we forget this feature of our epistemic practice, however, we will come to the conclusion that we cannot know the sceptical hypotheses not to obtain, since after all there is the uneliminated and ineliminable possibility that they might in fact obtain. Only if we keep in mind that being uneliminated is not enough for an error-possibility to be relevant can we see that we know very well that we are not brains in a vat (provided that we are not brains in a vat).[28]

[25] Cf. Wright 2002.

[26] Cf. Putnam 1998.

[27] Cf. Stroud 1989.

[28] Thanks to Alexander Bagattini, Hannes Ole Matthiessen, Andreas Maier, Patrick Leland and Shannon Hoff.

REFERENCES

Brandom, R. (1994). *Making It Explicit*, Cambridge, Mass.

Cappellen, H./LePore, E. (2005). *Insensitive Semantics: a Defense of Semantic Minimalism and Speech Act Pluralism*, Blackwell, Malden Mass.

Cohen, St. (1988). 'How to be a Fallibilist', *Philosophical Perspectives* 2, 91–123.

DeRose, K. (1995). 'Solving the Sceptical Problem', *Philosophical Review* 104, 1–52.

Goldman, A. I. (1976). 'Discrimination and Perceptual Knowledge', *Journal of Philosophy* 73, 771–791.

Lewis, D. (1996). 'Elusive Knowledge', *Australasian Journal of Philosophy* 74, 549–567.

Peirce, Ch. S. (1886). 'Some Consequences from Four Incapacities', *Journal of Speculative Philosophy* 2, 140–157.

Prichard, D. (2002). 'Two Forms of Epistemological Contextualism', *Grazer Philosophische Studien* 64, 19–55.

Putnam, H. (1998). 'Scepticism', *Philosophie in systematischer Absicht*, (ed.) M. Stamm, Klett-Cotta, Stuttgart, 239–268.

Schiffer, St. (1996). 'Contextualist Solutions to Scepticism', P*roceedings of the Aristotelian Society* 94, 317–333.

Stanley, J. (2004). 'On the Linguistic Basis for Contextualism', *Philosophical Studies* 119, 119–146.

Stroud, B. (1989). 'Understanding Human Knowledge in General', *Knowledge and Skepticism*, (eds.) M. Clay/K. Lehrer, Boulder, 31–49.

Williams, M. (1991). *Unnatural Doubts. Epistemological Realism and the Basis of Scepticism*, Oxford.

— (2001). *Problems of Knowledge*, Oxford.

— (2003). 'Skeptizismus und der Kontext der Philosophie', *Deutsche Zeitschrift für Philosophie* 51/6, 973–991.

Williamson, T. (2000). *Knowledge and Its Limits*, Oxford.

Wright, C. (2002). '(Anti-)Sceptics Simple and Subtle: G.E. Moore and John McDowell', *Philosophy and Phenomenological Research* Vol. LXV No.2, 330–348.

Grazer Philosophische Studien
74 (2007), 273–290.

EXISTENCE, INEXPRESSIBILITY AND PHILOSOPHICAL KNOWLEDGE

Dagfinn FØLLESDAL
Stanford University & University of Oslo

Summary

Ontology has traditionally been regarded as a core area of philosophy. However, during the 20th century, some philosophers have maintained that issues concerning existence and ontology are meaningless (Carnap) or inexpressible (Wittgenstein). Others, like Quine, have argued that these issues are both intelligible and important. After a short discussion of these views, the paper goes on to discuss the twist Husserl gives to our way of looking at this kind of philosophical knowledge through his notion of the thetic component of acts.

Existence has long been a major topic in philosophy; it is the central theme of ontology, which in turn is a main branch of metaphysics. Since metaphysics is a core area of philosophy, in the sense that whatever philosophical topic one deals with, one is bound to get into metaphysical issues, most philosophers have had something to say about metaphysics and in particular about ontology. However, much of what has been said about ontology and other metaphysical subjects has been rather murky, and it is no wonder that when Carnap and other logical empiricists tried to draw a line between the meaningful and the meaningless, metaphysics and ontology fell on the meaningless side of the divide.

Carnap did recognize that when he talked about there being physical objects, numbers, etc. this sounded like ontology. However, he claimed that there was no real issue here, it was all a matter of choosing the proper linguistic framework, but there was no claim that numbers and the like "really" existed, whatever that might mean.

Carnap was pressed into this position mainly by Quine, who queried him about the status of numbers and other abstract entities. Quine was not satisfied with Carnap's answer and his appeal to linguistic frameworks. For Quine, ontological issues were on a par with factual issues, and just as much subjects of scientific investigation as the factual issues. To ask what physical objects there are, is of the same ilk as asking what properties they have and what relations they bear to one another. Quine proposed his famous criterion of ontological commitment: "to be is to be the value of a variable". One believes in the existence of those objects that belong to the universe of discourse of the theories one believes in, whether they are theories of natural science or mathematical theories. Questions of what the world is like include questions of what objects there are in the world as well as what properties these objects have and what relations they bear to one another.

It is for this reason that Quine could not be a nominalist. The article "Steps toward a Constructive Nominalism" (1947), which Quine wrote together with Goodman, starts with the sentence "We do not believe in abstract entities". This was for Goodman a basic tenet of his world view. He could not get himself to believe in abstract entities. When they found in their article that very little of mathematics could be made sense of without quantifying over numbers and other abstract objects, Goodman concluded "so much the worse for mathematics" while Quine took the article to be a *reductio ad absurdum* of the attempt to be a nominalist. In order to be a nominalist one had to give up not only mathematics, but also almost all of natural science, building as it does on mathematics that goes far beyond the nominalist's meager ontology.

My original interest in philosophy was very much oriented towards ontological issues. I was frustrated with the classical attempts to deal with these questions. I found them very far from clear. In some cases I sensed that they were on to something important, but it was hard to tell what it was. I regarded Thomas Aquinas's arguments for the existence of God with great scepticism (and my scepticism also applies to Gödel's argument (Gödel 1970), which I read much later). However, Aquinas's reflections on existence seemed to me to be on to something, although it was difficult to pinpoint exactly what it was. Etienne Gilson seemed to me to be on to the same when 700 years later he wrote that the existence of a dot of moss was sufficient to convince him of the existence of God.

Inexpressibility

Wittgenstein in the *Tractatus* took another tack. According to him, "Not *how* the world is is the mystical, but *that* it is". This would be a way of avoiding the challenge of saying something about the issue of existence, but still hint that there was something there, and something important, as was also the case with ethics, religion and much else.

A first question I want to say something about is this notion of there being something inexpressible, something that cannot be expressed in language. Wittgenstein is only one of very many philosophers who have claimed that there are some matters that are inexpressible. While I was preparing this lecture I came across a recent book by André Kukla, *Ineffability and Philosophy* (2005), that surveys a variety of such views and systematizes them, first by what it is that is supposed to be inexpressible (facts—whatever this may be, or something expressible in *mentalese*, but not in other languages, or something else), secondly according to what kind of language this is supposed to be inexpressible in (some formal language or other, some natural language, all natural languages, etc.), and various other factors. The book then consists in a listing and brief discussion of all combinations of these features. Regrettably, the discussions of the various combinations are very brief. Kukla also, unfortunately, uses notions like "mentalese" very uncritically, without seeming to be aware of how unclear and problematic these notions are.

In some cases one can give clear and interesting notions of inexpressibility. Thus, for example, in the case of formal languages, what is expressible in one language may not be expressible in another, and interesting questions arise about expressibility and non-expressibility, questions that often have a bearing on questions of decidability and completeness. Tarski gives interesting examples of this, and one of the first papers I wrote was comments on a paper on "Expressive completeness" at the APA, Eastern Division meeting in New York in 1962.

However, in the case of natural languages the issues of inexpressibility are much harder to get a hold on. What is it that cannot be expressed? Facts? What are facts? One common answer is that facts are composits of objects with various properties and with various relations holding between them. Then, if we have names for the objects and words for the various properties and relations that are involved, we should be able to express these facts. In cases where some of the objects are nameless or where we have no words for some of the properties and relations that are involved, we can perhaps say that this fact is inexpressible in our language, although there may be indirect ways of expressing them. One can even prove that for a given language there are innumerable facts that are inexpressible in this sense. Thus, for example, as Cantor showed through his diagonal argument, there are nondenumerably many real numbers. Normal languages contain only a denumerable number of expressions, and hence there will not be enough numerals to go around to all the numbers. However, this does not mean that there are real numbers that have no name in any language whatsoever. For any given real number one could construct a language that has a name for that number. Given the freedom we have to construct names in a natural language it seems that for most purposes we do not have to be concerned about the fact that the real numbers outnumber the expressions in a language.

There are many interesting questions connected with indirect ways of expressing facts. Thus, we can often use quantifiers instead of names to express the existence of objects, without having to worry about whether the object has a name. For example, if we prove that there is a unique object having a specific property, is not this existential statement expressing a fact about that object without the need to introduce a name for the object? Of course, given its uniqueness, we can introduce a definite description for the object, but we can assert the existence of an object without having to introduce a description of the object or give it some name.

Wittgenstein's *Tractatus* doctrine of the limitations imposed upon us by language is beset by difficulties and hard to defend, and he did eventually give it up. In the *Philosophical Investigations* and other works he discusses instead the many ways in which we use language and also goes into substantive issues, such as that of ultimate justification, in *On Certainty*. One might hope that he would then have something to say on the issues that in the *Tractatus* he regarded as so important, but also so intractable that one had to be silent about them, such as questions of existence or of ethics. However, he remains remarkably silent about them.

Quine was the first I read who considered these issues to be important issues of philosophy and also succeeded in saying something intelligible about them. He was remarkable for his ability to think and write clearly about issues I regarded as "deep" and important. Of him it can truly be said that he was one of the few intelligent creatures at home in deep waters, to use a quip from Daniel Dennett's humorous little *Philosophical Lexicon*.

It was reading Quine that made me decide to give up mathematics and science in favor of philosophy. I had long had an interest in philosophy. However, I could not envision having philosophy as a job. So I settled for science, with philosophy as hobby—until I read Quine.

Quine gave a clear and precise account of how we determine what there is: We work out our best scientific theory and find out what objects are quantified over. The term "exists" is a general term. However, we do not as in the case of other general terms start by determining the term's meaning and then ascertaining its extension. Quine does not tell us what the meaning of the term "exists" is. He would agree with Frege that it is a second-level concept under which first-level concepts fall. However, like Frege, he does not go beyond this. Morton White, in his book *Toward Reunion in Philosophy* (1956), takes up the issue of the meaning of "exists" and concentrates on the question whether the word "exists" has several senses, for example one for abstract entities and on for concrete ones, as has been argued by various philosophers, many of whom have even used different terms for the existence of objects of different kinds, for example "being" for all objects and "existence" for spatio-temporal ones. White argues that "exists" is unambiguous. There are clearly differences between the objects in question; to claim that there is in addition a difference in the way they exist is to introduce a distinction that is neither called for nor supported by arguments.

Husserl

Husserl is a philosopher who has some interesting thoughts on this issue. Since this is a neglected topic even among Husserl scholars, I will use this opportunity to say something about it. Husserl discusses this question under the heading the thetic component of consciousness.

Husserl starts off from Brentano's theory of intentionality. According to Brentano, our consciousness is characterized by intentionality. Our consciousness is structured into *acts*, which are characterized by being directed towards objects. Brentano characterizes this directedness in the following oft-quoted passage:

> Every mental phenomenon is characterized by what the Scholastics of the Middle Ages called the intentional (or mental) inexistence of an object, and what we might call, though not wholly unambiguously, reference to a content, direction toward an object (which is not to be understood here as meaning a thing), or immanent objectivity. Every mental phenomenon includes something as object within itself, although they do not do so in the same way. In presentation, something is presented, in judgment something is affirmed or denied, in love loved, in hate hated, in desire desired and so on.
>
> This intentional inexistence is characteristic exclusively of mental phenomena. No physical phenomenon exhibits anything like it. We can, therefore, define

> mental phenomena by saying that they are those phenomena which contain an object intentionally within themselves.[1]

While Brentano's examples seem convincing, other cases raise problems. What when a person hallucinates? Is there an object towards which an act of hallucination is directed? And what about people who believed in Pegasus or who were trying to calculate the largest prime number? Did their acts have objects?

Brentano struggled with these problems till the end of his life. His students proposed different solutions. Alexius Meinong insisted that all acts have an object, but that this object does not always exist. There are existing objects, but there are also non-existing ones. Pegasus is one of the non-existing ones, so is the largest prime number. Brentano did not appreciate this rescue attempt. He could not make sense of existence as a property, which some objects have, others not.

Husserl proposed another solution. He maintained that acts need not have an object, but they have directedness. They are always *as if* directed towards an object, but this does not guarantee that they have an object. The challenge is to clarify what this directedness consists in. Husserl's phenomenology can be characterized as an attempt to meet this challenge.

Acts

In our normal lives we are absorbed by the world and its objects or we are engaging in other forms of activity. Husserl calls all these activities *acts*. Many acts involve movements of our bodies. Others are intellectual or emotional. They, too, involve physiological processes in our organisms and may be prompted by or lead to bodily acts. They all involve our consciousness.

The phenomenological reduction

In 1905 Husserl got the idea of the phenomenological reduction, which for him is intimately connected with his idealism. The phenomenological reduction starts from our natural, world-directed attitude. Instead of attending to the world and its objects we bracket the objects in the world

[1] Brentano, *Psychology From an Empirical Standpoint*: 50 of Terrell's translation.

and are not concerned with them and their existence. Instead, we are focusing on the acts. The aim of phenomenology is to study, in detail, the structures of acts.

Noema, noesis, hyle

In focusing on the acts, we discover *three* elements: the *noema*, the *noesis* and the *hyle*. The *noema* is a meaning, a structure, which interrelates all the features of consciousness that go into the act. The noema is an attempt to capture what Husserl sometimes calls the manner of givenness of an object; he is interested in the correlation between experienced object and its various manners of givenness.[2] In a first approximation, the noema could be compared to a set of anticipations: all the anticipations we have with regard to the object. The anticipations relate to all the different features of the object, the ones which already "meet the eye" and those which we have not yet observed. Most of these anticipations are not thematized, this is one reason that the word "anticipation" is not quite appropriate. However, it may be pedagogically helpful, to give us an idea of what Husserl is after.

The noema has several components, one of which is the thetic component, which will be the topic of this paper. However, before we turn to it, let us note that the noema has no temporal coordinates. It contains determinations of the temporal features of the objects of acts, but it is not itself temporal. It can, in principle, be the same in several acts, acts of the same agent that take place at different times or even acts carried out by different agents (although so much of the agent's peculiarities and background and of the spatio-temporal setting is involved in the noema that in practice no two agents would ever have the same noema, and even for one agent to have the same noema twice would be problematic).

The notion of the noema may help us make the notion of an act a little clearer. We can individuate acts by saying that an act comprises all the components of consciousness that are unified by a noema. Acts are the same only if they have the same noema. Note the "only if", as we noted, two acts can in principle have the same noema. Since the noema has no

[2] In the *Crisis* Husserl emphasizes the importance of this topic: "The first breakthrough of this universal a priori of correlation between experienced object and manners of givenness (which occurred during work on my Logical Investigations around 1898) affected me so deeply that my whole subsequent life-work has been dominated by the task of systematically elaborating on this a priori of correlation" (*Krisis*, Husserliana, VI, 169, n. 1 = 166 of Carr's translation).

temporal coordinates it is not a part of the act. The act has temporal coordinates and can only have parts that are temporal.

The other two elements that we discover when we carry out the reduction and reflect on the act, the noesis and the hyle, are, however, temporal and are parts of the act. In fact, together they make up the act, acts have no other parts. The noesis and the hyle are *experiences* that we have; they have duration in time, in a special sense that we are not going to discuss here. All acts have a *noesis*. This is a very special kind of experience, which gives meaning, or structure to the act. Husserl calls the noesis the meaning-giving element of the act, and the noema he calls the meaning given in the act. As one should expect, there is a thorough-going parallelism between noema and noesis. An example Husserl gives in order to clarify the two notions is that of a judgment. Philosophers since Bolzano have argued that what we study in logic, are abstract entities called judgments, and not the acts through which we make the judgments. The former are components in noemata, while the latter include the noesis.

The third element on our list, the *hyle*, are experiences which we typically have when our sense organs are affected, but we also can have in special other situations, for example when we are affected by fever, drugs or nervous disturbances. They form a kind of boundary condition for the kind of noesis we can have in acts of perception. For perception to take place, the noesis and the hyle must fit harmoniously together. Note that we may keep our eyes open and think about something else, for example a philosophical or a mathematical problem. In such a case we may have hyle, but the hyle do not play any role in determining the object of our act. The thetic character of the act is not that of perception, but that of thinking. We do not perceive.

A central point in Husserl's theory of perception, that we shall not discuss here, is that the noesis is never uniquely determined by the hyle. We can have very different noeses, and perceive very different objects, while what reaches our sensory organs may be the same. One should, however, not say that the hyle are the same in such a case. The hyle are not one-to-one correlated with the impingements on our sensory surfaces. The hyle are experiences, and not only is the noesis dependent on the hyle, also the hyle will depend on the noesis. There are no hyle that can be compared from act to act where the noeses are different. The important points for Husserl are that perception is underdetermined by what reaches our sensory organs, and that there is nothing given in perception. Perception is directly of objects, and there are no intermediary steps. Neither the hyle, nor the noesis nor the noema are objects that we perceive. The former two are experiences, not objects experienced (except, of course when we are turning them into objects of study in the phenomenological reduction). And the noema includes a structure that the perceived object has, but it also includes much more. And it is not the physical object that we perceive.

The thetic element in consciousness

With this as a background, let us now turn to the thetic element in consciousness. As always when we deal with consciousness, we may approach it through noema or through noesis. We shall here concentrate on the noema, but what is said can, *mutatis mutandis*, be put into a noetic framework.

Let us then consider a noema. The noema has, basically, two parts, a noematic sense, or meaning, and a thetic component. The sense, or meaning, contains components corresponding to the properties of the object, and it also includes the idea, through the so-called "determinable *X*" that an object is not a collection of properties, but something that has properties. Further, the determinable *X* reflects the idea that two objects may be very similar and still not identical, while one and the same object may display quite different properties at different times or when seen from different points of view.[3]

We shall not go into these elements of the noema, but concentrate on the thetic component. Let us begin with *perception*. If we compare an act of perception and an act of remembering, these acts may have the same object and hence have much in common as far as the noematic sense, or meaning, goes. However, they are very different acts, and this is reflected in the thetic component of the noema (as well as in the noesis and hyle). In the case of perception, the thetic component involves a number of different elements that have to fit together: First, in acts of perception, the *hyle* play a role determining, in interplay with the noesis, what object we see, hear, smell or feel. Thereby they also have an influence on the meaning-component of the noema. This concerns not only the hyle we have now. We also have anticipations concerning the various hyle that we will have if we move around, follow the object through time, etc., and the act also has to fit in with the hyle we had earlier in our experience of the object.

[3] A discussion of the determinable *X* may be found in Follesdal 2001.

The presence of hyle is not enough to make an act an act of perception. As we noted earlier, we can be in the same sensory situation, with our eyes open, etc, but think of something else. In the latter case, the object of our act may be an abstract entity, or we may think of a person or an object far away. In such a case, we are not perceiving. One characteristic of the thetic component in perception is hence that in perception the hyle play a role: the noesis and hyle must harmonize with one another.

Remembering is different. We may remember an object we once perceived, but the hyle we now have, will normally be irrelevant to the act of remembering. In some cases, the hyle we have may be relevant, there may be something about the present situation that reminds me of the object. Also the object I remember may be likely to have left traces, which I may now look for and that may help corroborate what I remember or make it less plausible. There is hence a connection between memory and present sensation that constrains us, and this gives to memory, as it gives to perception, a reality-character. What is remembered, and what is perceived, is experienced as real. Remembering and perceiving are here unlike *fantasizing,* which is unencumbered by my hyle. I can fantasize whatever I want. I may fantasize that there is a horse in this room, but I cannot perceive a horse now, however hard I try. Unlike fantasy, perception is not up to us, neither is memory. The prize we have to pay for the freedom of fantasy is that what we fantasize is not real. The recalcitrance that is present in perception and in memory plays an important part in giving these kinds of acts their reality-character.

Husserl was very interested in the thetic component of consciousness, the features of consciousness that distinguish different kinds of acts, such as acts of perceiving, acts of remembering, acts of imagining, etc. In particular, Husserl was interested in the difference between acts where we experience things as real and acts in which what we experience has a different status, for example, is experienced as merely imagined or dreamt.

The theme was brought up in the *Logical Investigations*,[4] but after Husserl's conversion to idealism and introduction of the phenomenologi-

[4] In the *Ideas* Husserl writes: "In the *Logische Untersuchungen* they [the posited moments] were (under the title 'quality') taken into the concept of sense (of significational essence) and therefore in this unity the two components, 'matter' (sense, in the present conception) and quality, were distinguished. [Here Husserl refers in a footnote to *Logische Untersuchungen*, V, §§20–21, Findlay's English translation: 586–593.] But it seems more suitable to define the term 'sense' as merely that 'matter' and then to designate the unity of sense and thetic character as '*positum* [*Satz*]'". *Ideen,* Husserliana, III.1,305.8–15 = 274 of the original edition = 317 of Kersten's translation, which I have slightly amended.

cal reduction it took on a new form. All questions of existence and reality are bracketed in the phenomenological reduction. That they are bracketed does, however, not mean that they are gone. They are there, but we are no longer asking what is real and what is not real. Instead we ask: What is involved in being real? What are the structures of consciousness thanks to which we experience something as real? And how do they differ from the structures of our consciousness when we experience something as dream or fantasy? Indeed, one of the central concerns of Husserl's idealism is to get an understanding of how the being of the world and its objects is represented in our consciousness. We will return to his idealism later.

Reality character and perception

One of Husserl's chief concerns was to understand better this reality-character, which he saw as intimately connected with perception:

> To answer these questions I shall look for the ultimate source which feeds the general positing of the world effected by me in the natural attitude, the source which, therefore, makes it possible that I consciously find a factually existing world of physical things confronting me and that I can ascribe to myself in this world and am able to assign myself a place there. Obviously, this ultimate source is *sensuous experience*. For our purposes, however, it will be sufficient if we consider *sensuous perception* …[5]

So, for Husserl, the sensuous experiences, the hyle, are a key to our understanding how it comes that we understand the world as real.

There is a similarity here between Husserl and William James, who writes: "Sensible vividness or pungency is then the vital factor in reality …".[6] There may have been some influence here of James on Husserl, who became aware of James *Principles* in the early 1890s through his teacher Stumpf. However, while James has only short remarks on this topic, Husserl explored and developed it rather fully. Especially in his later works he emphasized the role that the body and our bodily activity play for our conception of reality.

[5] *Ideen*, Husserliana, III.1,80.33–81.1 = 70 of the original edition = 82 of Kersten's translation.

[6] *Principles of Psychology*, Ch. XXI, Vol. 2: 930 of the Harvard edition, 1983.

The lifeworld

Note how Husserl in the passage above states that we posit the world in the natural attitude. We study this positing in the phenomenological attitude, where we study our acts and their structures. However, the acts we study are acts that we carry out in the natural attitude. In other passages Husserl makes clear that this positing of the world that we perform in the natural attitude is not a *judgment* that the world exists:

> The general positing … *does not consist of a particular act*, perchance an articulated judgment *about* existence. It is, after all, something that lasts continuously throughout the whole duration of the attitude, i.e., throughout natural waking life. … in short, everything which is, before any thinking, an object of experiential consciousness … bears … the characteristic "there", "on hand"; and it is essentially possible to base on this characteristic an explicit (predicative) judgment of existence agreeing with it. If we state such a judgment, we nevertheless know that in it we have only made thematic and conceived as a predicate what already was somehow inherent, as unthematic, unthought, unpredicated, in the original experiencing or, correlatively, in the experienced, as the characteristic of something "on hand".[7]

This is a theme to which Husserl often reverts, that our consciousness is mostly not thematized. What is thematized is only the tip of an iceberg. The rest is hidden to us, but makes itself known when we encounter "recalcitrant experience", which makes us aware of anticipations we never had thought of. This is a theme that has been taken up by Michael Polanyi (Polanyi 1958) and many others, and by them has been called "tacit knowledge". However, Husserl studied this characteristic of consciousness a generation earlier, and with his usual thoroughness, he noticed many features that have been overlooked by later authors. Thus, for example, Polanyi's phrase "tacit knowledge" suggests that the unthematized parts of our consciousness have propositional form, like knowledge, or are at least of one of the two forms "knowing that" or ”knowing how". However, Husserl observed that it is not these kind of advanced structures that are "the ultimate source which feeds the general positing of the world", but our sensuous experiences. Before we come to explicit judgments, or even to unthematized propositions, we therefore have a long way to go.

Husserl called this rich and complex structure of largely unthematized

[7] *Ideen*, Husserliana, III.1,62.1–17 = 53 of the original edition = 57–58 of Kersten's translation.

consciousness, or rather the world that would correspond to it, "the lifeworld". A key observation he made, which I regard as an intriguing contribution to our contemporary discussion of ultimate justification, is that every claim to validity and truth rests upon this "iceberg" of unthematized prejudgmental acceptances.[8] One should think that this would make things even worse for justification. Not only do we fall back on something that is uncertain, but on something that we have not even thought about, and have therefore never subjected to conscious testing. Husserl argues, however, that it is just the unthematized nature of the lifeworld that makes it the ultimate ground of justification. "Acceptance" and "belief" are not attitudes that we *decide* to have through any act of judicative decision. What we accept, and the phenomenon of acceptance itself, are integral to our lifeworld, and there is no way of starting from scratch. Only the lifeworld can be an ultimate court of appeal:

> Thus alone can that ultimate understanding of the world be attained, behind which, since it is ultimate, there is nothing more that can be sensefully inquired for, nothing more to understand.[9]

The existence of the world is, according to Husserl, indubitable. He writes:

> … the lifeworld, for us who wakingly live in it, is always there, existing in advance for us, the "ground" of all praxis, whether theoretical or extratheoretical. The world is pregiven to us, the waking, always somehow practically interested subjects, not occasionally but always and necessarily as the universal field of all actual and possible praxis, as horizon. To live is always to live-in-certainty-of-the-world.[10]

Husserl's idealism does hence not consist in rejecting the reality of the world, or regarding it as an illusion. On the contrary, the very notion of an illusion presupposes the reality-character of the world. To say that the world is an illusion would verge on a contradiction. It would be to undercut the very sense of what is claimed. There are certain parallels to this in the earlier German idealists, notably Fichte. However, Husserl's position seems to me to be better thought through, and it differs in important respects from the positions that are commonly labeled "idealism". Husserl

[8] For more on this, see Follesdal 1988.

[9] *Formale und transzendentale Logik*, § 96b, Husserliana XVII, 249.18–20 = Cairn's translation: 242.

[10] *Krisis*, § 37, Husserliana VI, 145.24–32 = Carr's translation: 142.

was notorious for his lack of skill in understanding other philosophers and for his ineptitude in using their terms. Thus, for example, terms like "ontology" and "metaphysics" are used in a very idiosyncratic way by Husserl. My own view is that the traditional idealism/realism distinction is ill suited to capture Husserl's position, and that here, as in the rest of his philosophy, he might have been better off avoiding traditional philosophic terminology. This is confirmed by a letter he wrote in 1934 to Abbé Baudin: "No ordinary 'realist' has ever been as realistic and concrete as I, the phenomenological 'idealist' (a word which by the way I no longer use)".[11]

In the Preface to the first English edition of the *Ideas* (1931), Husserl stated:

> Phenomenological idealism does not deny the factual [*wirklich*] existence of the real [*real*] world (and in the first instance nature) as if it deemed it an illusion … Its only task and accomplishment is to clarify the sense [*Sinn*] of this world, just that sense in which we all regard it as really existing and as really valid. That the world exists … is quite indubitable. Another matter is to understand this indubitability which is the basis for life and science and clarify the basis for its claim.[12]

Intersubjectivity

Husserl emphasizes that the lifeworld is an intersubjective world. We do not conceive of the world in which we live as a private world, not accessible by others. On the contrary, we regard the world as a shared world, a world that we all experience, although from different perspectives. Husserl emphasizes this in his discussions of the lifeworld, and he also stresses it in connection with his discussion of the thetic component of our acts:

> I take their surrounding world and mine Objectively as one and the same world of which we are all conscious, only in different modes …
> For all that, we come to an understanding with our fellow human being[s] and in common with them posit an Objective spatiotemporal actuality as *our factually existent surrounding world to which we ourselves nonetheless belong* …
> As what confronts me, I continually find the one spatiotemporal actuality

[11] Letter quoted in Iso Kern: 276.

[12] Husserl, Preface to the Gibson's translation of *Ideas*, Allen & Unwin, London, 1931. Here from the German version in Husserliana, V,152.32–153.5, my translation.

> to which I belong like all other human beings who are to be found in it and are related to it as I am.[13]

Note that in this quotation from 1913, Husserl does not use the phrase "lifeworld", but instead talks about the "surrounding world". While the idea of the lifeworld comes up early in his work, the word "lifeworld" occurs for the first time in a manuscript from 1917, and it first was used in print in *Crisis of the European Sciences*, a small part of which was published in the journal *Philosophia*, Belgrade, in 1936, but the important discussion of the lifeworld did not come out until 1976, when the full manuscript, *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, was published as Volume 6 of the Husserliana, the standard edition of Husserl's works.[14]

Other kinds of thetic components

The thetic component in acts of perception is particularly interesting, since it is connected with our conceptions of reality and existence. The thetic component of acts of remembering also has such a connection, although, as we noted, in a more roundabout way. Husserl therefore often uses perceptual acts as examples, because they are among the simplest acts: they are *thetically one-membered*, their thetic component does not involve reference to acts with other thetic components.

Most other acts are further removed from the reality-positing typical of perception. However, Husserl notes that

> We will find grounds for extending the concept of positing to all act-spheres and thus speak of, e.g., liking-positing, wishing-positing, willing-positing, with their noematic correlates "liked", "wished-for", "ought to be in the practical realm", and the like.[15]

[13] *Ideen*, Husserliana, III.1,60.16–18 and 24–26, and 61.15–18 = 52 of the original edition = 55–57 of Kersten's translation.

[14] See Follesdal 1990.

[15] *Ideen*, Husserliana, III.1,60.16–18 and 24–26, and 61.15–18 = 234 of the original edition = 270 of Kersten's translation, slightly amended.

Values

We will not here go into Husserl's analyses of all the different kinds of positing. I will, however, end this paper by discussing one observation that Husserl makes, and that I find rather intriguing. Husserl writes:

> The new sense brings in a totally *new dimension of sense*; with it no new determining parts of mere "things" are constituted, but instead *values of things*, value-qualities, or concrete Objects with values: beauty and ugliness, goodness and badness; the use-Object, the art work, the machine, the book, the action, the deed, and so forth.[16]

Husserl hence gives to values a status similar to that of objects and their properties. He had a conception of secondary qualities according to which they are just as much a part of the objective world as are the primary qualities, and values have a similar status. Presumably, values have all the features that we have discussed above: our constitution of values is subject to constraints; we are not, as Sartre held, free to create our values any way we want, but experience constraints that constrict us in a way similar to that in which the hyle constrain us in what we are able to perceive. The values that are construed subject to these constraints are experienced as intersubjective, valid for all. They are therefore not merely expressions of an individual's likes and dislikes, which make no claim on others. To consider an example of distributive justice: if we consider only ourselves we might prefer that a cake be so divided that we get the larger piece, but we are constrained in the direction of a conception of distributive justice and may come to regard a division into equal pieces as just. We thereby pass from subjective preferences to objective values. Husserl, in his manuscripts on ethics, some of which have now appeared in print[17], discusses how our subjective likes and dislikes get turned into objective values through adjustments involving symmetry between persons, etc.

This brings us to the intersubjectivity of values. Not only the things in the world and their properties are conceived of as shared and intersubjective, so are also values. We may disagree on values, but we do not treat them merely as likes or dislikes and say that *de gustibus non est disputandum*. Instead, we discuss them, and we think that we may be right or wrong about them. We also think that we can argue about them and give evidence

[16] *Ideen*, Husserliana, III.1,267.5–10 = 239–240 of the original edition = 277 of Kersten's translation, slightly amended.

[17] Roth 1960, and Edmund Husserl, *Vorlesungen über Ethik und Wertlehre 1908–1914*.

for some views and against others. This may be generally accepted now, after Rawls, but in the years that separate Husserl and Rawls this has not been a very popular view.

REFERENCES

Brentano, F. (1874): *Psychology From an Empirical Standpoint,* 1874, Meiner, Hamburg, 1924 and later editions. English translation by D. B. Terrell of Volume 1, Book 2, Chapter 1, (ed.) Roderick B. Chisholm, *Realism and the Background of Phenomenology*, Free Press, Glencoe, Ill., 1960.

Dennett, D. (1987). *The Philosophical Lexicon*, edited with K. Lambert, 8th Edition, published by the American Philosophical Association 1987.

Føllesdal, D. (1988). 'Husserl on evidence and justification', *Edmund Husserl and the Phenomenological Tradition: Essays in Phenomenology (Proceedings of a lecture series in the Fall of 1985.)*, (ed.) Robert Sokolowski, Studies in Philosophy and the History of Philosophy, Vol. 18, The Catholic University of America Press, Washington, 1988, 107–129.

— (1990). 'The Lebenswelt in Husserl', *Language, Knowledge, and Intentionality: Perspectives on the Philosophy of Jaakko Hintikka* (Acta Philosophica Fennica, Vol. 49), (eds.) Leila Haaparanta, Martin Kusch, and Ilkka Niiniluoto, Helsinki, 1990, 123–143.

— (2001). 'Bolzano, Frege and Husserl on Reference and Object', *Future Pasts: The Analytic Tradition in Twentieth-Century Philosophy*, (eds.) J. Floyd and S. Shieh, Oxford University Press, Oxford, 2001, 67–80.

Gödel, K. (1970). 'Ontological proof', (1970), *Kurt Gödel, Collected Works, Vol. III: Unpublished Essays and Lectures*, (eds.) S. Feferman et. al., Oxford University Press, Oxford, 1995, 403–404.

Goodman, N., and Quine, W. V. O. (1947). 'Steps toward a Constructive Nominalism', *Journal of Symbolic Logic* 12, 105–122.

Husserl, E. (1900–01). *Logische Untersuchungen. Zweiter Teil. Untersuchungen zur Phänomenologie und Theorie der Erkenntnis. In zwei Bänden*, U. Panzer (ed.) (Husserliana, Vol. XIX), Martinus Nijhoff, The Hague, 1984. English translation by J. N. Findlay, *Logical investigations*, Routledge and K. Paul, London, 1970.

— (1913). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Erstes Buch: Allgemeine Einführung in die reine Phänomenologie. 1. Halbband: Text der 1.–3. Auflage*, K. Schuhmann (ed.) (Husserliana, Vol. III/1), M. Nijhoff, The Hague, 1976. English translation by W. R. Boyce Gibson, *Ideas: general introduction to pure phenomenology*, Macmillan, New York, 1931. English translation by F. Kersten, M. Nijhoff, The Hague, 1982.

— (1929). *Formale und transcendentale Logik. Versuch einer Kritik der logischen Vernunft*, P. Janssen (ed.) (Husserliana, Vol. XVII), Martinus Nijhoff, The Hague, 1974. English translation by D. Cairns, *Formal and transcendental logic*, Martinus Nijhoff, The Hague, 1969.

— (1954). *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie. Eine Einleitung in die phänomenologische Philosophie*, W. Biemel (ed.) (Husserliana, Vol. VI), Martinus Nijhoff, The Hague, 1954. English translation by D. Carr, *The crisis of European sciences and transcendental phenomenology. An introduction to phenomenological philosophy*, Northwestern University Press, Evanston, IL, 1970.

— (1971). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie. Drittes Buch: Die Phänomenologie und die Fundamente der Wissenschaften*, W. Biemel (ed.) (Husserliana, Vol. V), Martinus Nijhoff, The Hague, 1971. English translation by T.E. Klein and W.E. Pohl, *Ideas pertaining to a pure phenomenology and to a phenomenological philosophy. Third Book. Phenomenology and the foundations of the sciences*, M. Nijhoff Publishers, The Hague, 1980.

— (1988). *Vorlesungen über Ethik und Wertlehre 1908–1914*, U. Melle (ed.) (Husserliana, XXVIII), Kluwer, The Hague, 1988.

James, W. (1990). *The Principles of Psychology*, 2 volumes, Macmillan, London, 1890, Harvard University Press, Cambridge, Mass., 1983.

Kern, I. (1964). *Husserl und Kant. Eine Untersuchung über Husserls Verhältnis zu Kant und zum Neukantianismus* (Phenomenologica 16), Martinus Nijhoff, The Hague, 1964.

Kukla, A. (2005). *Ineffability and Philosophy*, Routledge, London, 2005.

Polanyi, M. (1958). *Personal Knowledge*, Routledge, London, 1958.

Roth, A. (1960). *Edmund Husserls ethische Untersuchungen: dargestellt anhand seiner Vorlesungsmanuskripte* (Phenomenologica 7), Martinus Nijhoff, The Hague, 1960.

White, M. (1956). *Toward Reunion in Philosophy*, Harvard University Press, Cambridge, Mass.

Grazer Philosophische Studien
74 (2007), 291–305.

CONTEXTUALISM AND THE BACKGROUND OF (PHILOSOPHICAL) JUSTIFICATION

Christian BEYER
Universität Erfurt

Summary

I propose to apply a version of contextualism about knowledge to the special case that represents the topic of this volume. I begin by motivating my preferred version of contextualism, which may be labelled as conventionalist contextualism; here I start from a well-known problem that besets epistemic internalism (section I). Following this, I pose a problem for conventionalist contextualism and argue that it can be solved by invoking, first, the idea of what I shall call the lifewordly background of epistemic justification, an idea originating from Husserl and Wittgenstein, and, secondly, the associated notion of normality, to be found in Husserl (section II). Finally, I apply the resulting conception of justification to philosophical knowledge, particularly focussing on the special role of intuitions (section III).

I.

The (access-)internalist conception of justification has it that in order for an epistemic subject to be justified in a given belief, he (or she)[1] must be able to make explicit (to himself or others)—i.e. must have cognitive access to—his justifying reasons for holding that belief; with these reasons being themselves part of the subject's overall system of beliefs. Externalism rejects this accessibility requirement, claiming that external relations such as the reliability of the respective belief-forming mechanism may be sufficient for epistemic justification. One of the disadvantages of this view is that it does not seem to do enough justice to the idea that justification goes hand in hand with epistemic and practical responsibility, both of which appear to require that the subject be capable of critically reflecting upon his justifying reasons (but see fn. 11 below). Thus, a clairvoyant may be

[1] In what follows, I shall drop this addition.

"justified" in the externalist sense without being responsible either in his mode of belief-acquisition or in his according practical decisions; indeed, this appears to hold true even if the clairvoyant lacks cognitive access to any anti-reliability reason (cf. BonJour 2003: 31f). However, internalism faces its own difficulties, the most serious one being *the epistemic regress problem*: the beliefs you need to be able to make explicit in order to be justified must themselves be justified, which according to internalism means that you need to be able to make explicit your reasons for these justifying beliefs, where those reasons have in turn to be cognitively accessible, and so on, *in infinitum*. So it looks like we can never be conclusively justified in any belief whatsoever.

In order to stop this unpleasant regress, internalists such as Chisholm have developed foundationalist conceptions of justification, according to which some beliefs—the ones that provide your ultimate justification for all other beliefs—are directly justified, i.e. their justification does not depend on other justified members of your belief-system, to which they are inferentially related. According to Chisholm, these "basic" beliefs are *cogito*-like. Thus, if you are to justify your observational belief that (1) the mountain peak is white, you may invoke your belief that (2) it appears to you as if you were just seeing a white mountain peak, and this latter belief will be directly justified.

However, as Sellars has pointed out, a *Konstatierung* like (2) ("It appears to me as if I were seeing a white mountain peak") can justify an observational belief such as (1) only if the respective subject is also justified in both his belief that (3) normally, i.e. in standard conditions, a belief such as (2) is caused by the presence of a white mountain peak and his belief that (4) the relevant standard conditions obtain (cf. Sellars 1997: 74f). But this means that the regress has not come to a halt after all. For, beliefs like (3) and (4) stand again in need of justification. I shall call this *the normality problem*, and I will return to it later.

To solve the epistemic regress problem in a more satisfactory way, coherentist conceptions of justification have been proposed. According to coherentism, beliefs are justified holistically rather than in a linear manner. That is to say: all of your justified beliefs mutually support each other by being part of a coherent, and thus (in the primary sense) justified, belief-system. Hence, we get a non-malicious circle rather than a malicious regress. However, this view is beset with many difficulties, having partly to do with the need to account for the essential role of observation when it comes to justifying empirical beliefs, and partly with the internalist requirement that the coherence of a belief-system be cognitively accessible in order for any given member of that system to be justified holistically.

Let us take stock. Both epistemic externalism, (internalist) foundationalism and (internalist) coherentism all have their own merits but display serious disadvantages as well. It would be nice if we had a conception of justification at our disposal which preserves those merits but avoids the respective disadvantages. Here epistemic contextualism fills the bill (cf. Brendel 2001: 101–105). As an additional advantage, it may help us meet the sceptical challenge.

Quite generally, contextualism holds that a given belief's epistemic justification is dependent upon the context in (or relative to) which the belief is assessed. For, acccording to contextualism, whether or not a subject is justified in a given belief, in a context of epistemic assessment *c*, depends on which *criteria* or *standards for justification* are relevant in *c*. Therefore, sentences of the type "*S* is justified in his belief that *p*" are claimed to be context-sensitive, such that they are true in some contexts of assessment but false in others.

This approach allows for a rather traditional definition of knowledge that is compatible both with internalism (for some contexts of assessment) and with externalism (for other contexts)—depending on how we "deformalize" its third condition:

> *Contextualism about knowledge*
> *S* knows that *p* at time *t* in a context of assessment *c* iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* meets, at *t*, the criteria (standards) of justification relevant in *c*.

Accordingly, contextualism (about knowledge) claims that sentences of the type "S knows that *p*" are context-sensitive, where their truth-value is a function of the relevant context of assessment with its associated criteria of justification. These criteria can in turn be looked upon as determining the *relevant alternatives* (i.e., possible worlds) the subject must be in a position to rule out (as non-actual) if his belief is to count as a case of knowledge in the relevant context (cf. Pryor 2001: 97f).

What is it, though, that determines the criteria of justification relevant in a given context? There would seem to be at least four options: (i) the knowledge-attributor; (ii) the subject whose belief is epistemically assessed; (iii) the linguistic community, or culture, the knowledge-attributor belongs to; (iv) the linguistic community, or culture, the subject in question belongs to. The last two options are generally overlooked.[2] In what follows, I shall focus on (i) and (iii). If we choose option (i), we arrive at a view that has been aptly called *radical contextualism* (cf. Ernst 2005). On this view, it is the attributor alone who sets the standards for justification and knowledge. According to Ernst, radical contextualism thus holds that:

> *Radical contextualism (version #1)*
> *S* knows that *p* at time *t* for a given speaker making a corresponding knowledge ascription iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* is able, at *t*, to rule out the alternatives taken into account by the speaker.[3]

However, this view seems to me to lead into an inacceptable relativism, according to which knowledge lies solely in the eye of the beholder. For, on this view, say, a farmer who believes, on the basis of his perception of a red sunset, that the weather will be fine, may (provided it *will* be fine) at the same qualify as *knowing* and *being ignorant* about the weather for a given speaker, *full stop,* depending solely on which alternatives the speaker happens to take into account, i.e., which context of assessment, with its according standards for justification, he chooses to regard as relevant (presumably when making his knowledge ascription). If this is not simply a contradiction, it comes close to being one: the farmer counts as being ignorant about the weather (for the speaker) in a possible world that merely differs from a neighbouring world where he counts as knowing about it (for that speaker) by being such that in this world the speaker regards another set of alternatives as relevant. Moreover, it implies that there is no difference between someone's *being* a reliable informant about the weather and his being *held* to be one (pace Ernst 2005: 170ff). To my mind, these unpalatable consequences can only be avoided by relativizing the *definiendum* (and not merely, like in the foregoing version, the definiens) of the proposed definition to a relevant assessment context (with associated relevant alternatives), thus:

[2] For an overview on the current spectrum of contextualist positions see e.g. Blaauw 2005: I–XVI, esp. sec. 1.

[3] Cf. Ernst 2005: 164. The sentence scheme "S knows that p" is mentioned, rather than used, in Ernst's definiendum, but if we relativize its being satisfied to a speaker (as I think we have to, so as to avoid another inacceptable relativism, resulting from the fact that different speakers may employ different criteria of justification), his definition is equivalent to the one under consideration (except that I have added a time-index).

> *Radical contextualism (version #2)*
> *S* knows that *p* at time *t* for a given speaker in a context of assessment *c* iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* meets, at *t*, the criteria of justification that, according to the speaker, apply in *c*.
> (In other words:)
> *S* knows that *p* at time *t* for a given speaker in a context of assessment *c* iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* is able, at *t*, to rule out the alternatives taken into account by the speaker in (or relative to) *c*.

This relativization blocks the undesired relativism, since even if we consider the immediate neighbourhood of a given possible world, "The farmer knows about the weather in context c_1 (for speaker *A*)" is clearly compatible with "The farmer is ignorant about the weather in context c_2 (for *A*)".

However, the present view still has the (to my mind) problematic consequence that our farmer may be a reliable informant about the weather (in a given context) *for me*, whilst he fails to be one (in that context) *for somebody else* who—unlike me—rejects, as unjustified, beliefs based on peasant's weather maxims (cf. Baumann 2002: 80f). It seems to be a better idea to delegate this epistemic decision either to the *majority* or (even better) to the *recognized experts* from our linguistic community—say, to meteorologists who are free of self-conceit. Therefore, it appears preferable to go in for the more moderate option (iii) above:

> *Conventionalist contextualism*
> *S* knows that *p* at time *t* for a speaker belonging to a given linguistic community *l*, in a context of assessment *c*, iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* meets, at *t*, the criteria of justification that, according to the conventions valid in *l*, apply in *c*.
> (In other words:)
> *S* knows that *p* at time *t* for a given linguistic community *l* in a context of assessment *c* iff (1) *S* believes at *t* that *p*, (2) it is true that *p*, and (3) *S* is able, at *t*, to rule out the alternatives to be taken into account by the speaker in *c* according to the conventions valid in *l*.

Like other versions of contextualism, this view allows for a decent reply to the sceptic who claims that we do not know we are not a brain in a vat located near Alpha Centauri. The reply is that (a) in a sceptical context—say: in a seminar on scepticism—we do fail to know this all right, nor do our ordinary empirical beliefs amount to knowledge here; but (b) in a given everyday context we still know we are not brains in a vat, provided we know, by normal standards, some ordinary empirical proposition which we know to entail the former proposition. For, there is mutual (implicit) agreement between all of us to the effect that in a context such as this less strict standards for justification apply. Note that we do not have to let the principle of deductive closure (of knowledge) go by the board in order to reach this conclusion.

Furthermore, the present view enables us to make a reasonable compromise on the externalism/internalism issue. After all, we sometimes favour internalist (e.g. coherentist) standards for justification, but then again sometimes we rather favour externalist (e.g. reliabilist) criteria. Compare our knowledge ascriptions regarding, on the one hand, a mathematician concerning, say, a complex equation and, on the other hand, a small child of whom we want to say that he knows by perception that there is food on the table. Conventionalist contextualism accommodates these varying intuitions by stressing the context-dependence of conventional standards for justification.

What about the notorious Gettier problem? It seems to me that the present account is able to cope with it. So consider Smith, who inferentially forms a true belief to the effect that *p* (the successful applicant has 10 coins in his pocket/Jones owns a Ford or Brown is in Barcelona/one of his colleagues owns a Ford) on the basis of his belief that *q* (Jones will get the job and has 10 coins in his pocket/Jones owns a Ford/some—from his viewpoint—trustworthy colleague told him that he owns a Ford), where the belief that *q* is supposed to be false, or at least misleading (for Smith), but justified—so that Smith fails to know that *p* while allegedly still being justified.

It is the last assumption that conventionalist contextualism can, and should, challenge. The respective Gettier-type example is bound to be described in such a way that the description makes it plausible to the competent reader to assume that *q* can be false, or misleading (for Smith), in the described sort of situation, while *p* is true. Therefore, we are dealing with a case where according to the conventions governing its description the uneliminated alternative that *q* is false, or misleading, is to be taken into account by the attributor, i.e. by the addressee of the example (who should be regarded as belonging to the same linguistic community as the person bringing forward the example).[4]

[4] A similar point is made by Ernst 2002. Ernst stresses that Gettier-type examples direct

Thus, for instance, in the example where, irrespective of some evidence of Smith's to the contrary, Jones does not own a Ford (but Brown happens to be in Barcelona), Smith fails to rule out an alternative as non-actual that the attributor cannot but regard as relevant, as the example is described in a way that makes it clear to him that in this kind of situation it may indeed be actual—notably, the possibility that Jones does not really own a Ford. Or again, in the example where Nogot, the colleague of Smith's who has so far always told him the truth, is lying about his alleged possession of a Ford, the uneliminated possibility that Smith's evidence is defective is bound to become relevant for the example's addressee. By conventionalist contextualism, Smith fails to know that *p* for that addressee, i.e. for ourselves, because his belief that *p* (although correctly derived) is based upon a belief which is unjustified for us in the context of assessment created by the description of the example.

To put the same point in a different manner: the very description of any given Gettier-type example creates an assessment-context such that the person bringing forward the example as a counter-example against conventionalist contextualism, and any addressee following him in this regard, perform a pragmatic contradiction. The way an example such as this is designed, its protagonist (i.e. Smith) fails to rule out a possibility that he would have to be able to eliminate in order to meet the criteria of justification that, according to the conventions of their linguistic community, apply in the context of assessment created by the description of the example; yet it is claimed that he is indeed justified.

II.

One epistemological problem remains, however. (*At least* one problem, that is.) *What justifies the relevant criteria of justification?* It is at this point that the notion of the lifewordly background enters the scene.

"*Lifeworld*" is a label introduced by Husserl to denote the way the members of one or more social groups (cultures, linguistic communities) use to structure the world into objects (cf. Husserl 1970: sec. 34f, 36, pp. 132, 138; henceforth *Crisis*). The respective lifeworld is claimed to "predelineate" a "world-horizon" of potential future experiences that are to be

our attention to various "possibilities of error" and thus "suggest to us" to look upon them from the perspective of a "doubter" (cf. ibid: 126). He does not bring in conventionalist contextualism to explain this datum, though.

(more or less) expected for a given group member at a given time, under various conditions, with the resulting sequences of anticipated experiences corresponding to different possible worlds. (Here the contextualist notion of relevant alternatives fits in.) These expectations follow typical patterns, as the lifeworld is fixed by a system of (first and foremost implicit) conventions that determine what counts as "normal" or "standard" observation under "normal" conditions.[5] Some of these conventions are restricted to a particular culture,[6] whereas others determine a "general structure" that is "a priori" in being "unconditionally valid for all subjects", defining "that on which normal Europeans, normal Hindus, Chinese, etc., agree in spite of all relativity" (cf. *Crisis*: sec. 36, p. 139). Husserl quotes universally accepted facts about "spatial shape, motion, sense-quality" as well as our prescientific notions of "spatiotemporality", "body" and "causality" as examples (Cf. ibid.). These conceptions determine the general structure of all particular thing-concepts that are such that any creature sharing our typically human epistemic constitution will be capable of forming and grasping them, respectively, under different lifeworldly conditions. If you will, it is this universal "a priori" structure of the lifeworld that makes intercultural understanding possible (cf. *Husserliana* XV: text #11, p. 159).

"*Background*" is Wittgenstein's term for a subject's "*Weltbild* (*picture of the world*)", i.e. for the system of criteria on the basis of which the subject "distinguishes between true and false" (cf. Wittgenstein *On Certainty*: sec. 94). I propose to think of these criteria as standards for justification that are determined by the subject's lifeworld. This fits in well with Wittgenstein's view that there are "Moorean" beliefs which simply cannot be doubted as long as our inherited *Weltbild* does not yield any reason for calling them into question. The—usually unthematized—beliefs expressed by the respective sentences "I use to live near the surface of the Earth" (Wittgenstein's example) and "No one can walk through walls" are cases in point. Wittgenstein characterizes the according statements as "hardened empirical propositions (*erstarrte Erfahrungssätze*)" (cf. ibid.: sec. 96) that belong to the basic rules of the game of giving and asking for reasons. If you violate a rule such as this, in that you retain, say, an abnormal belief such as "I

[5] For the relationship between the notions of "lifeworld" and "normality" cf., e.g., Husserl 1973 (henceforth *Husserliana* XV): text #10, pp. 135ff. For the importance of social conventions in this connection see, e.g., ibid.: 142.

[6] Cf., e.g., *Husserliana* XV: text #10, pp. 141f. Husserl also sometimes speaks of "relative normalities" in this context, which normalities are restricted to particular "homeworlds" (cf. ibid.: suppl. XIII, pp. 227-236).

am a ghost without a body", then you automatically count as unjustified. Your knowledge claim flies in the face of your fellow's intuitions about what experience normally teaches us (or rather: about how we use to live our lives)[7]. All you can do in this sort of situation is to try to establish your statement as a new rule of the game, for instance, by convincing your fellow epistemic subjects that it is after all possible to walk through walls. To achieve this, you have to fight their fixed intuitions; you must try to bring about strong conflicting intuitions, so as to make them realize that their system of accepted empirical propositions is actually incoherent.

In a situation such as this, a fundamental epistemic change may occur. We are dealing with a "fluid empirical proposition (*flüssiger Erfahrungssatz*)" that may become a new "hardened empirical proposition" and then sink into a new background of what is generally taken for granted. As a consequence, the lifeworld of the corresponding group of epistemic subjects will change. For example, the world may become re-structured in such a way that their former notion of a human being gets replaced by a successor notion which applies to human-like ghosts as well.[8]

My conjecture is that when things get thus into flux, coherentist criteria of justification become predominant,[9] whereas foundationalist and externalist criteria play a particularly significant role against the background of a fixed lifeworld. However, there will always be some more foundationalist elements in play, even when one lifeworld gets replaced by another, notably what I have been calling intuitions; where these experiences can (as a first approximation; but see below) be looked upon as spontanous, involuntary acceptances manifesting usually unthematized, i.e. implicit, but rather firm beliefs. On the other hand, it is the lifewordly background that generally determines our intuitions. When our epistemic foundations shake, we get confused by conflicting intuitions, some of which already predelineate the

[7] Cf. *Crisis*: sec. 37, p. 142: "[T]he life-world, for us who wakingly live in it, is always already there …, the 'ground' of all praxis whether theoretical or extratheoretical. The world is pregiven to us … as the universal field of all actual and possible praxis, as horizon. To live is always to live-in-certainty-of-the-world".

[8] Husserl observes that the hardened empirical propositions (and hence the lifewordly horizon) can vary between different cultures at a given time: "We have a world-horizon as a horizon of possible thing-experience [*Dingerfahrung*] …; but everything is here subjective and relative, even though normally, in our experience and in the social group united with us in the community of life, we arrive at 'secure' facts … But when we are thrown into an alien social sphere, that of the Negroes in the Congo, Chinese peasants, etc., we discover that their truths, the facts that are for them fixed, generally verified or verifiable, are by no means the same as ours" (*Crisis*: sec. 36, pp. 138f).

[9] Cf., e.g., *Crisis*: sec. 38, p. 145; cf. also Føllesdal 1988: 117ff, 121f.

structure of the forthcoming lifeworld. It partly depends on us—on both the coherentist criteria we choose to apply and the intuitions we choose to take seriously—, what this new lifeworld will look like.

For Husserl, this process of re-structuring the world involves mutual agreements among epistemic subjects, i.e. it is (partly) a matter of convention.[10] So the resulting standards for justification will be just as rational as the underlying conventions. But now the tricky question arises whether there may be a sense in which a convention can be *epistemically* rational or justified (in terms of how well it helps us achieve our epistemic goals, i.e. to believe the truth and to avoid error), as opposed to being merely *practically* justified (in terms of how well it helps us satisfy our ordinary desires).[11] Surely, some conventions may help us solve social coordination problems such as the prisoners' dilemma. But the question remains whether a merit such as this may be invoked to justify a given convention epistemically, should the corresponding need arise.

Anyway, to take up a picture that is often used to illustrate epistemic foundationalism: after a lifeworldly change such as this, new trees of justification can grow, some of whose branches are going to end with what Wittgenstein calls "hardened empirical propositions". Other branches will, however, rather end with (introspectively) conscious instances of belief types (like perception- or memory-based belief) whose normally unconditional acceptance, under what intuitively counts, against the respective lifewordly background, as standard conditions, also belongs to the rules of the game. Thus Husserl observes:

> The life-world is a realm of original self-evidences. That which is self-evidently given is, in perception, experienced as 'the thing itself', in immediate presence, or in memory, remembered as the thing itself … All conceivable verification leads back to these modes of self-evidence, because the 'thing itself' (in the particular mode) lies in these intuitions themselves as that which is

[10] For a clear expression of this view, cf. Husserl 1984: sec. 51, pp. 23ff, where Husserl refers to what he later was to call the intersubjective lifeworld under the label "communicative environment".

[11] Cf. Sosa 2003: 102f: "An athlete may be helped to win by her strong and steady confidence that she *will* win, which may provide her with practical justification for somehow acquiring and sustaining that confidence even in the teeth of contrary evidence. But such practical justification does not bear on whether she knows what she believes, unlike the evidence against her belief". This consideration makes it clear that epistemic justification is not logically necessary for practical justification, which, however, still leaves open the possibility that there are many contexts (such as the one described in Bonjour 2003) in which the latter presupposes the former.

> actually, intersubjectively experiencable and verifiable … (*Crisis*: sec. 34d, pp. 127f)

I hasten to add that Husserl stresses again and again our notorious fallibility as experiencing subjects. Accordingly, he allows for changes in what generally counts as "intersubjectively experiencable and verifiable".

Let us apply this admittedly rather sketchy view to the normality problem. Against a fixed lifeworldly background, it makes no sense to ask for a justification of beliefs regarding standard conditions, because these beliefs, and the sorts of circumstances in which they are relevant, are built right into the conventions that guide us, against that background, in playing the game of giving and asking for reasons. So here the regress of justification ends with our corresponding intuitive acceptances. However, if things get into flux, if a change of lifeworld appears necessary, in view of strong recalcitrant intuitions, then these conventions (and even the principles of logic)[12] turn out to be by no means sacrosanct; we can then meaningfully ask for an intersubjective justification of them. Thus, it may become debatable whether there are ghosts after all, or, to cite a perhaps more realistic example, whether clairvoyance can still be accepted as an ultimate source of justification.

III.

Where does this leave us with respect to philosophical knowledge? When a philosopher claims that a given example exemplifies a property that he is interested in (say: knowledge) (cf. Goldman and Pust 1998: 182) or when he makes a general statement that he regards as necessarily true (cf. Sosa 1998: 258), then he will, no doubt, typically appeal to intuitions. However, an appeal to intuitions such as this is to be received with caution; the intuitions in question, call them philosophical intuitions, are often less than secure. This has to do with the typical mode of philosophical inquiry. For, more than any ordinary discipline, philosophy seems to be in a permanent state of flux. New intuitions are caused and tested by means of thought experiments, by counterfactual variations on recent scientific hypotheses that may (just like the theoretical identification of water with H_2O-stuff) have an impact on the prescientific lifeworld, and so on. More

[12] Cf. *Crisis*: sec. 34, p. 135, as well as Husserl's critique of the Vienna circle in sec. 35, p. 141. Cf. also Føllesdal 1988: 123–128.

or less reluctantly, new systems of thought are proposed to integrate *some* of these intuitions; since we are fallible beings, there will always be conflicting intuitions, anyway (cf. Sosa 1998: 261). Paradoxes such as Theseus' Ship are familiar cases in point.

Although some paradoxes can surely be resolved without abandoning any underlying intuition, we should, absent a satisfactory phenomenology of intuition, refrain from classifying all philosophical intuitions as (manifestations of) beliefs; some of them are perhaps merely *inclinations* to believe, or conscious states in which a belief-disposition such as this manifests itself (cf. Sosa 1998: 258f).

Indeed, one of the tasks of philosophical inquiry is precisely to investigate into the structure of possible lifeworlds that are constrained by certain of these intuitions. Think of more recent philosophical debates about personal identity, for instance (headwords: teleportation, irreversible brain splits). These debates can be conceived of as intuition-driven inquiries into ways the world could, or should, be structured, or re-structured, if certain engineering procedures turned out to become practicable, which would be bound to change our everyday lifeworld to some extend.

More conservative philosophers, call them descriptive metaphysicians, confine themselves to making explicit the intuitive beliefs associated with our present lifeworld and integrating them within a coherent conceptual network that is supposed to represent that lifeworld.

With yet other philosophers (myself included) it is somewhat unclear just how conservative their investigations may turn out to be. They try to integrate more recent scientific hypotheses within their picture of the world, starting from the assumption that these theories may be used to make explicit, or explicate, and indeed to justify so far unthematized background beliefs pertaining to our lifeworld without too much incoherence thus coming to the foreground. Where those background beliefs can in turn be employed to justify the relevant hypotheses holistically, but also to modify these hypotheses if necessary.

The underlying working assumption is a Husserlian one: despite its notorious subject- and culture-dependence, it is our common lifeworld, and the associated intuitive acceptances, that provide the "grounding soil" (cf. *Crisis*: sec. 34e, p. 131) of the more objective world of science; notably in the twofold sense that (i) scientific conceptions owe their (sub-)propositional content and thus their reference to reality to the prescientific notions they are supposed to "naturalize" and that, consequently, (ii) when things get into flux in science, when a *crisis* occurs, all that is left to appeal to in order to defend new scientific approaches against their rivals is our system of prescientific background beliefs, as manifested in our associated intuitions.[13] Thus, to quote two recent examples where people's intuitions seem to differ widely, if we are to decide between theory theory and simulation theory about attitude ascription, or between higher-order and first-order representation views on consciousness, it may be a good idea to consult our lifewordly background and see how well these different theories fit in there. However, in order to do so, we first need to make explicit both these theories and the implicit beliefs constituting the relevant region of our (prescientific) lifeworld. If we find ourselves disclosing incoherencies in our belief-system this way, appropriate scientific hypotheses can be used to repair these leaks in Neurath's ship.

A philosophical finding such as this is the more interesting since well-established scientific conceptions may indeed change the structure of the common lifeworld.[14] Thus, our intuitive acceptances are sometimes rather theory-loaded—whether this be a good thing for philosophical purposes (cf. Kornblith 1998: 132–137) or not. In any case, I think it is part of our job as philosophers to make conceptual offers such as this, in terms of how the world might be re-structured to get a more coherent lifewordly background. It is in this sense that the task of philosophy can be characterized as conceptual basic research.

By conventionalist contextualism, it follows that the extent to which philosophical knowledge deviates from what is generally taken for granted—or how conservative it turns out to be—ultimately depends on the lifewordly background which is presupposed by the standards for justification that count as relevant in the respective philosophical context of assessment. Thus, revisionary metaphysics creates different assessment contexts than descriptive metaphysics or more or less conservative naturalistic approaches. As a consequence, the standards to be invoked to justify philosophical claims concerning lifeworldly structures will vary. These standards are constrained by intuitions about actual and counterfactual cases. I have suggested that epistemologically these intuitions are in the same boat as perception, memory and explicit "Moorean" belief. As Sosa

[13] See Føllesdal 1990: 139f, ## 2 and 3, and the references cited to support these points.

[14] The claim that scientific conceptions my change the structure of our everyday lifeword should be kept apart from the Husserlian thesis that in a sense the scienctific world represents a *part* of the intersubjective lifeworld (i.e. point # 1 in Føllesdal 1990). The latter thesis seems to amount to the claim that "the adults of our time" always already conceive of the world "as being in principle scientifically determinable" (Husserl 1973: sec. 10, p. 42).

observes, some of them are particularly “obdurate”: “thus, perhaps, the intuition that it is possible for one to know that one sees a … hand” (cf. Sosa 2003: 159). In the light of the foregoing considerations, it seems promising to proceed on the following assumption in this regard. The closer to the culture-independent “general structure” of the lifeworld (see above) the area of the lifeworld associated with the respective intuition is located—in other words: the greater the probability that the intuition is shared by all normal humans—, the harder it is to abandon that intuition. To the extent that epistemology reflects on obdurate intuitions such as this, it should therefore be regarded as constrained by general anthropology.

REFERENCES

Baumann, Peter (2002). *Erkenntnistheorie*, Metzler, Stuttgart/Weimar.

Blaauw, Martin (2005). ‘Introduction: Epistemological Contextualism’, *Grazer Philosophische Studien* 69, I–XVI.

Bonjour, Laurence. (2003). ‘A Version of Internalist Foundationalism’, *Epistemic Justification*, (eds.) L. Bonjour and E. Sosa, Blackwell, Oxford, 5–96.

Brendel, Elke. (2001). ‘Eine kontextualistische Lösung des Streits zwischen Internalisten und Externalisten in der Erkenntnistheorie’, *Erkenntnistheorie*, T. Grundmann (ed.), mentis, Paderborn, 90–106.

Ernst, Gerhard (2002). *Das Problem des Wissens*, mentis, Paderborn.

— (2005). ‘Radikaler Kontextualismus’, *Zeitschrift für Philosophische Forschung* 59, 159–178.

Føllesdal, Dagfinn. (1988). ‘Husserl on Evidence and Justification’, *Edmund Husserl and the Phenomenological Tradition*, R. Sokolowski (ed.), The Catholic University of America Press, Washington, 107–129.

— (1990). ‘The *Lebenswelt* in Husserl’, *Language, Knowledge and Intentionality* (Acta Philosophica Fennica 49), L. Haaparanta et al. (eds.), Helsinki, 123–143.

Goldman, Alvin and Pust, Joel. (1998). ‘Philosophical Theory and Intuitional Evidence’, *Rethinking Intuition*, M. DePaul and W. Ramsey (eds.), Rowman & Littlefield, London, 179–197.

Husserl, Edmund (1970). *The Crisis of European Sciences and Transcendental Phenomenology* [*Crisis*], transl. by David Carr, Northwestern UP, Evanston.

— (1973). *Experience and Judgement*, transl. by Churchill and Ameriks, Northwestern UP, Evanston.

Husserl, Edmund (1973). *Zur Phänomenologie der Intersubjektivität Dritter Teil: 1929–1935* (= *Husserliana* XV) [*Husserliana* XV], ed. by Iso Kern, Nijhoff, The Hague.

— (1984). *Die Konstitution der geistigen Welt* [from *Ideas II*], Meiner, Hamburg.

Kornblith, Hilary. (1998). 'The Role of Intuition in Philosophical Inquiry', *Rethinking Intuition*, M. DePaul and W. Ramsey (eds.), Rowman & Littlefield, London, 129–141.

Pryor, James (2001). 'Highlights of Recent Epistemology', *Brit. J. Phil. Sci.* 52, 95–124.

Sellars, Wilfrid (1997). *Empiricism and the Philosophy of Mind*, with an introd. by R. Rorty and a study guide by R. Brandom, Harvard UP, Cambridge/Mass.

Sosa, Ernest. (1998). 'Minimal Intuition', *Rethinking Intuition*, M. DePaul and W. Ramsey (eds.), Rowman & Littlefield, London, 257–269.

— (2003). 'Beyond Internal Foundations to External Values', in L. Bonjour and E. Sosa: *Epistemic Justification*, Blackwell, Oxford, 97–170.

Wittgenstein, Ludwig. *On Certainty*.