Transcendental Philosophy and Naturalism

edited by
JOEL SMITH AND
PETER SULLIVAN

Transcendental Philosophy and Naturalism

This page intentionally left blank

Transcendental Philosophy and Naturalism

EDITED BY

Joel Smith and Peter Sullivan

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS
Great Clarendon Street, Oxford OX2 6DP
Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in
Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto
With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam
Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries
Published in the United States
by Oxford University Press Inc., New York

First published 2011

British Library Cataloguing in Publication Data
Data available
Library of Congress Cataloging in Publication Data
Data available
Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Books Group, Bodmin and King's Lynn
ISBN 978–0–19–960855–3
1 3 5 7 9 10 8 6 4 2

Contents

Acknowledgements

This is one of two volumes originating in the AHRC-funded project *Transcendental Philosophy and Naturalism* (2005–2008). The papers in this volume were presented at various workshops and conferences organized by the project in Essex, London, Oxford, and Cambridge. We gratefully acknowledge the support of the AHRC, and would like to thank all of the other institutions and people who helped to make this project a success.

The project was conceived and directed by Mark Sacks (1953–2008). Mark's untimely death has been a great loss to everyone who knew him. This volume is dedicated to his memory.

JS, PMS

Contributors

Adrian Haddock, University of Stirling
Patricia Kitcher, Columbia University
Hilary Kornblith, University of Massachusetts
Penelope Maddy, University of California, Irvine
A. W. Moore, University of Oxford
Joel Smith, University of Manchester
Robert Stern, University of Sheffield
Peter Sullivan, University of Stirling

This page intentionally left blank

1

Introduction: Transcendental Philosophy and Naturalism

Joel Smith and Peter Sullivan

What is the contemporary status of transcendental philosophy? Is the relationship between transcendental philosophy and naturalism necessarily antagonistic? Does transcendental philosophy pose a significant challenge to contemporary naturalism? Can the naturalist offer a plausible debunking of the claims of transcendental philosophy? These are among the questions addressed by the chapters in this volume. In this introduction we will offer an overview of some prominent strands within the transcendental tradition and of their relation to some varieties of contemporary philosophical naturalism.

1. Transcendental philosophy

With the publication of the *Critique of Pure Reason*, and its introduction of distinctively transcendental forms of explanation and argument, and the associated metaphysical framework of transcendental idealism, Kant inaugurated a new philosophical tradition. While it will inevitably be from Kant that we derive our conception of what constitutes transcendental philosophy, subsequent work within this tradition has taken on a wide variety of forms, not all of which are self-consciously Kantian: philosophers as diverse as Fichte, Hegel, Schopenhauer, Nietzsche, Husserl, Heidegger, Carnap, Wittgenstein, Strawson, and Davidson can all reasonably be

placed, more or less centrally, within it.[1] Given this variety, an attempt to offer necessary and sufficient conditions for qualifying as a transcendental philosopher would surely be fruitless. Nonetheless it does seem reasonable to suppose that there a number of recognizable marks at least some of which will be displayed by those philosophical positions and approaches to philosophical problems that will count as, broadly speaking, within the transcendental tradition. Adopting at first a relatively narrow conception one might say that there are at least two commitments which would surely qualify someone as a transcendental philosopher. The first, doctrinal commitment would be endorsement of some form of transcendental idealism. The second, methodological commitment would be the production and deployment of transcendental arguments. So we will need to have before us some elementary account of both transcendental idealism and transcendental forms of argument. But in offering this account we will aim to be sensitive to a broader conception, according to which participation in a philosophical tradition is marked less by doctrine than by aspiration. It is perhaps particularly true of Kant's successors that they are united more by a shared sense of the central problems of philosophy, and of the kind of illumination to be expected from a resolution of them, than by commitment to the theoretical framework in which Kant's own solutions were developed. It has indeed been characteristic of the tradition to explore ways of recovering Kant's insights while surrendering or actively contesting the most ambitious aspects of his own constructive theorizing. Kant himself formulated—but did not stick to—the advice that certain of his notions are suited only to a 'negative use': for the broader conception of the tradition we have in mind it will be engagement with these notions, rather than endorsement of them, that is decisive.

1.1 Transcendental idealism

A good place to begin is with Kant's well-known remark in the Fourth Paralogism:

> I understand by the **transcendental idealism** of all appearances the doctrine that they are all together to be regarded as mere representations and not things in themselves, and accordingly that space and time are only sensible forms of our intuition, but not determinations given for themselves or conditions of objects as things in themselves. To this idealism is opposed **transcendental realism**, which

[1] See the various papers in (Gardner and Grist Forthcoming).

regards space and time as something given in themselves (independent of our sensibility). The transcendental realist therefore represents outer appearances (if their reality is conceded) as things in themselves, which would exist independently of us and our sensibility and thus would also be outside us according to pure concepts of the understanding . . . The transcendental idealist . . . can be an empirical realist . . . For because he allows this matter and even its inner possibility to be valid only for appearance—which, separated from our sensibility, is nothing—matter for him is only a species of representations (intuition), which are called external, not as if they related to objects that are **external in themselves** but because they relate perceptions to space, where all things are external to one another, but that space itself is in us. (Kant 1781/7: A369–70)[2]

On this picture, the transcendental realist (mistakenly, according to Kant) identifies appearances with things in themselves. The transcendental idealist, on the other hand, distinguishes these, thereby refusing to treat space and time as properties of things in themselves. Crucially, the transcendental idealist distinguishes between two levels of explanation, the empirical and the transcendental. So far as concerns 'outer appearances' the *empirically real* is simply that which is in space, the *empirically ideal* is that which is within the mind. This distinction is to be kept apart from that between the *transcendentally real*, which is that which is independent of the *a priori* necessary conditions of cognition (our sensibility), and the *transcendentally ideal*, which is that which is necessarily subject to such conditions. These divisions allow us to give a straightforward explanation of the Kantian claim that the transcendental idealist holds appearances to be empirically real yet transcendentally ideal. Appearances are in space rather than in the mind, but are nevertheless necessarily subject to the necessary conditions of cognition. Thus, the transcendental idealist is, at the same time, an empirical realist.

Of course, this minimal characterization is the beginning not the end of an account of transcendental idealism. In particular, it is neutral between ontological and methodological readings of the doctrine.[3] But it is already clear that Kant implicitly links, via the notion of sensibility, transcendental idealism with claims concerning the *a priori* necessary conditions of the possibility of cognition. To say that appearances are transcendentally ideal

[2] Also see (Kant 1781/7: A373), (Allison 2004: 24), and (Franks 2005: 49–51).

[3] See (Guyer 1987), (Langton 1998), (Gardner 1999), (Van Cleve 1999), (Ameriks 2003), and (Allison 2004).

is to say that they are necessarily subject to the *a priori* conditions of cognition.[4] And this is the ground of Kant's 'supreme principle', that,

> The conditions of the **possibility of experience** in general are at the same time conditions of the **possibility of the objects of experience.** (Kant 1781/7: A158/B197)

According to the Kantian picture, it is transcendental idealism, operating through this principle, that allows us to see how synthetic *a priori* knowledge is possible. As Haddock, following Lear, puts it in his contribution to this volume, synthetic *a priori* knowledge is grounded in 'how we are minded', and knowledge so grounded can be knowledge of objects only insofar as our conception of the ontological independence of these objects is constrained by the requirement of their being objects of cognition to us. Thus, in Kant's account of the objects of cognition, explanatory priority is granted to requirements of our subjectivity (Bell 1999: 194–202). But the explanations offered by Kant's transcendental idealism come at a heavy price, which is that all cognition is limited by the 'boundaries of possible experience' (Kant 1781/7: Bxix). That is, given the distinction between appearances and things in themselves, whilst it is true that we can entertain thoughts about the latter, all human knowledge is limited to the former.[5]

That is Kant's own position, but it is not the only one that might be termed 'transcendental idealism'. In their contributions to this volume both Moore and Haddock discuss variants of transcendental idealism according to which the way things are in themselves is not even so much as thinkable. Such a contention represents one way in which a merely 'negative use' might be made of a notion given a positive role in Kant's own theorizing. A natural extension of it, developed in work by Sacks that is explored in Sullivan's contribution to this volume, is to hold that the very notion of an object's existing 'in itself' can have *no* role in a properly critical philosophy. When it is

[4] On Allison's reading this is what allows Kant to distinguish transcendental idealism from Berkeleian subjective idealism. For, while both appeal, in their accounts of unperceived entities and of actuality, to the notion of a possible perception, 'Berkeley's account of possible perception is essentially psychological in nature. To be possible means to be actually perceivable . . . In sharp contrast to this, Kant defines the possibility of perception in terms of the conformity to rules, that is, to *a priori* principles' (Allison 2004: 41). So, Kant's transcendental idealism can be distinguished from Berkeleian idealism because of Kant's insistence that there are *a priori* necessary conditions of the possibility of cognition. Also see (Gardner 1999: 42, 271–8).

[5] Sacks (2000: 199) takes this ignorance claim to be the central claim of Kantian transcendental idealism.

pressed to that extent—when it is, in Wittgenstein's phrase, 'strictly carried through' (Wittgenstein 1922: 5.64)—it is reasonable to question whether the resulting position is idealist only in its ancestry.

1.2 Transcendental arguments

The form, content, and ambitions of transcendental arguments have been much discussed over recent decades.[6] To offer some context for this discussion it is worth turning to the elucidation of the term 'transcendental' that Kant offers in the *Prolegomena*:

> the word 'transcendental' . . . does not signify something passing beyond all experience but something that indeed precedes it *a priori*, but that is intended simply to make cognition of experience possible. (Kant 1783: 373, n.48)[7]

The transcendental is that which makes cognition possible. We can call any claim to the effect that such and such is a condition of the possibility of cognition, a *transcendental claim*. This leaves open the possibility that there may also be transcendental claims that state conditions of the possibility of something more specific than cognition, for example experience, language, or self-consciousness.

Different philosophers use the phrase 'transcendental argument' in different ways. On one understanding, a transcendental argument is any argument that has a transcendental claim as its conclusion.[8] According to another understanding, perhaps the more common, a transcendental argument is an argument in the form of *modus ponens* in which a transcendental claim figures as the conditional premise.[9] On such a view, a

[6] See, in particular, the essays collected in (Stern 1999). Also see Sullivan's contribution to this volume.

[7] This is, of course, related to Kant's account of what a transcendental philosophy would be: 'I call all cognition transcendental that is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible *a priori*. A system of such concepts would be called transcendental philosophy' (Kant 1781/7: A11/B25). It is a nice question how Kant's conception of the methods and purpose of transcendental philosophy relate to subsequent conceptions in, for example, Fichte, Hegel, and Husserl. For discussion of such issues, see the various papers in (Gardner and Grist Forthcoming).

[8] 'Anything that might helpfully be called a transcendental argument should issue in some conditional to the effect that some conditioned would be impossible, if not for some condition' (Franks 2005: 204).

[9] 'Roughly, transcendental arguments are arguments of the form "There is experience; it is a condition of the possibility of experience that P; therefore, P." Kant sometimes substitutes "cognition" for experience, and other writers start from "conscious awareness" or "intelligible thought" instead' (Walker 2006: 238).

transcendental argument would state that *A* is the case, and that *B* is a condition of the possibility of *A*, and go on to conclude that *B* is the case.[10] Given the above characterization of transcendental claims, we would expect *A* to be some claim concerning the actuality of cognition, experience, language, or such like.

While it is potentially distracting this difference in usage is plainly of no importance. Significant points concerning the nature of transcendental arguments can be made using the notion of a transcendental claim, this notion being common to both understandings of them. For example, some will insist that transcendental arguments should provide some kind of argument against scepticism.[11] On such an account, it would be important that *A* be something that the sceptic cannot deny, and *B* something the sceptic typically does deny. A successful transcendental argument would then show the sceptic's position to be undermined by his commitments elsewhere. If, on the other hand, *A* were something that the sceptic is free to doubt, such as the claim that we have empirical knowledge, then a transcendental argument would have little anti-sceptical force. It is clear that this debate focuses on the content of the transcendental claim.

It is plausible that some transcendental arguments can be found in Kant's writings. Which of Kant's arguments are distinctively transcendental is more difficult to say. Candidates from the first *Critique* might include the Transcendental Deduction, discussed at length in Kitcher's contribution to this volume; the Refutation of Idealism, discussed in Maddy's contribution to this volume; and the Second Analogy. Each of these apparently starts from a premise stating the actuality of some form of cognition and argues that a necessary condition of this is the objective validity of the categories, outer experience or causality. For example, the Refutation of Idealism begins from the premise that, 'I am conscious of my existence as determined in time' (Kant 1781/7: B275), asserts the transcendental claim that, 'consciousness in time is necessarily combined with . . . the existence of the things outside me' (B276), concluding that this, 'proves the existence of

[10] Allison (1969: 227) complains that Strawson illegitimately moves from understanding A as 'experience or cognition', to understanding it as 'having a conception of experience or cognition'. Sacks (2005) argues that for such reasons we should understand transcendental arguments as involving something other than entailment relations between propositions. If true, this would affect the above characterization.

[11] For transcendental arguments as anti-sceptical, see (Sacks 2000: Ch. 8). For a view of (at least Kantian) transcendental arguments as lacking anti-sceptical intent, see (Franks 2005: Ch. 3).

objects in space outside me' (B275). What seems certain is that this argument contains a transcendental claim, to the effect that the existence of objects in space is a necessary condition of the possibility of my consciousness of my changing determinations over time, and that the core of the Refutation lies in the case that Kant presents for this claim.

Transcendental claims state conditions of the *possibility* of cognition, and this quite obviously involves a modal claim.[12] An initial formulation that would suit many transcendental claims would be: necessarily, if *A* then *B*. This certainly gets us the modal character, but transcendental claims are usually understood not only as necessary but also as universal. We might, then, think of transcendental claims as of the following form: necessarily, everything is such that if it is *A* then it is *B*. This will not fit all transcendental claims, but it will cover a good number of cases.[13] For example, it would appear to adequately characterize both Strawson's (1959) claim, discussed in Smith's contribution to this volume, that it is a necessary condition of being able to ascribe states of consciousness to oneself, that one be able to ascribe them to others, and Korsgaard's (1996) claim, discussed in Stern's contribution to this volume, that acting on the basis of a principle requires that one value one's own humanity.

Not only are transcendental claims standardly intended to be necessary and universal, they are usually thought of as *a priori* and, often, also as synthetic.[14] That transcendental claims are *a priori* is, for Kant at least, guaranteed by the fact that they are necessary and universal,[15] and also by the fact that they state conditions *of* experience and so cannot be derived *from* experience.

[12] We thus disagree with Pihlström's claim that, 'whatever necessity is involved in the transcendental argument is the necessity of the argument itself, of the relation between the premises and the conclusion' (Pihlström 2004: 292).

[13] It will not do for those transcendental claims that take the condition and conditioned to be (or hold of) distinct things. A second complication is that the conditional in a transcendental claim should almost certainly not be treated as the material conditional. For transcendental claims generally aspire to be explanatory in a way that the material condition is not. For instance, that 2 + 2 = 4 is a condition of the possibility of cognition. But I doubt that many would think of this as a transcendental claim.

[14] See (Sacks 2005).

[15] 'Experience . . . tells us, to be sure, what is, but never that it must necessarily be thus and not otherwise. For that very reason it gives us no true universality . . . Now, such universal cognitions, which at the same time have the character of inner necessity, must be clear and certain for themselves, independently of experience; hence one calls them a priori cognitions' (Kant 1781/7: A1–2).

The character of (a good many) transcendental claims as necessary, universal, synthetic, and *a priori* throws further light on the two ways of understanding transcendental arguments contrasted above. For as long as we offer anti-sceptical arguments that have synthetic *a priori* transcendental claims as premises—a transcendental argument in the second sense—then it will be incumbent upon us to provide an argument for that very transcendental claim—a transcendental argument in the first sense. In any actual case it will be obvious that such a 'backup' argument is needed, since we can hardly expect that the central claim of any interesting transcendental argument will be immediately evident. The more general point here is that this further argument will be needed to persuade a sceptic whose scepticism extends quite generally to the possibility of synthetic *a priori* knowledge.

What of the relation between transcendental arguments and transcendental idealism itself? This question is one that has received a great deal of discussion. Strawson (1959, 1966) notoriously proposed a number of anti-sceptical transcendental arguments that were intended to succeed without a reliance on transcendental idealism of any sort. Stroud (1968) influentially argued that without a commitment to transcendental idealism, or otherwise some form of verificationism, such arguments were bound to fail as they cannot plausibly cross the gap between appearance and reality. Partly in response to Stroud's critique, Stern (2000) and various authors in (Stern 1999) have argued in favour of a modest role for anti-sceptical transcendental arguments; one that sidesteps Stroudian concerns about the ambitious arguments that Strawson proposed.

There are at least two ways in which one might think that the success of a transcendental argument will depend on transcendental idealism. The first, hinted at above, is that transcendental idealism is, in part, an attempt to show how synthetic *a priori* judgement is possible. If transcendental arguments contain synthetic *a priori* claims as premises, some account needs to be given to show that we have a right to such claims. A second way in which transcendental idealism might be thought a necessary accompaniment to any successful transcendental argument is that its distinction between the empirically real and the transcendentally real might give us some sense of how a transcendental argument can cross the gap between appearance and reality (Sacks 2000: 274–5). To see this, consider again the Refutation of Idealism. One might argue that Kant has not done enough to show that for temporal experience to be possible there must actually *be*

spatial objects, only that it must *seem* that there are spatial objects.[16] Insofar as the appearance/reality distinction in play here tracks Kant's distinction between appearances and things in themselves, one might suppose that Kant ought to agree with this contention. However, if it is intended to track the distinction between empirical seeming and empirical reality, Kant's transcendental idealism might be thought to allow for a crossing of the gap. Kant *can* allow that a transcendental argument can move from claims about the mind, to claims about (empirical) reality, since the transcendental idealist holds that (empirical) reality is necessarily subject to the *a priori* conditions of cognition.

These two are only the most obvious motivations for linking transcendental arguments to transcendental idealism, so long as we adhere to Kant's own understanding of the latter. Clearly the different understandings of transcendental idealism mentioned in §1.1 will imply correspondingly different understandings of this relationship. There is nonetheless a consensus that the onus rests with contemporary defenders of transcendental arguments without transcendental idealism to show how their claims can be justified.[17]

1.3 Transcendental constraints and transcendental features

Sacks (2000: Ch. 6) makes a useful distinction between transcendental features and transcendental constraints. He writes,

> Roughly, a *transcendental constraint* indicates a dependence of empirical possibilities on a non-empirical structure, say, the structure of anything that can count as a mind. Such constraints will determine non-empirical limits of possible forms of experience . . . A merely *transcendental feature*, on the other hand, is significantly weaker. Transcendental features indicate the limitations implicitly determined by a range of available practices: a range comprising all those practices to which further alternatives cannot be made intelligible to those engaged in them . . . those transcendental features of what we can currently envisage are not constraints on what is possible. (Sacks 2000: 213)

Transcendental claims, as described above, are most naturally associated with transcendental constraints. Transcendental claims state necessary conditions of the possibility of cognition, so they will 'determine non-empirical limits of possible forms of experience'. Transcendental features, on the other hand, do

[16] Whether this is a reasonable criticism of Kant's argument is not the present concern.
[17] For just such a defence see (Peacocke 2009).

not circumscribe what is possible. They may appear to do so, but that appearance is merely a reflection of our current, contingent, inability to conceive of alternatives. Now, it would be possible to reformulate transcendental claims in terms of transcendental features rather than transcendental constraints. That is, one might claim that some aspect of cognition is (currently, contingently) inconceivable in the absence of some further thing. In the case of the example mentioned above, Kant's Refutation of Idealism, the claim would become that consciousness of my changing determinations over time is not (currently, contingently) conceivable without the existence of objects in space. This claim would be significantly weaker than the version of the claim stated in terms of transcendental constraints.

The distinction between transcendental constraints and transcendental features as we so far have it opens up the possibility of two distinct ways of understanding transcendental idealism. On the first, that employing transcendental constraints, transcendental idealism is the view that appearances are necessarily dependent on the necessary conditions of the possibility of cognition. On the second, that employing transcendental features, transcendental idealism would be the view that appearances are dependent on those features alternatives to which we cannot currently conceive; according to Sacks, it is something like this second picture that both Williams (1974) and Lear (1982, 1984) claim to find in the work of the later Wittgenstein.[18] But in fact transcendental features differ from transcendental constraints in a number of different ways—inconceivability replaces necessity, empirical determination replaces non-empirical determination, a contingent historical grounding replaces timeless and absolute universality, and so on—and Sacks' own arguments show that these several points of difference need not align. The contrast of constraints and features thus makes it possible to envisage formulating transcendental idealism in a number of intermediate ways, and these alternative formulations will then serve as an important tool for making further distinctions within the transcendental tradition.

2. Naturalistic philosophy

Philosophical naturalism comes in a number of forms. One can distinguish between ontological, methodological, and epistemological variants. The

[18] See also Haddock's chapter in this volume.

ontological naturalist makes a claim about the contents of the world, holding that only natural entities exist. The methodological naturalist makes a claim about our ways of coming to know about the world, holding that natural scientific methods of gaining knowledge are in some way privileged. The epistemological naturalist makes a claim about the nature of knowledge, holding that knowledge is a natural phenomenon to be investigated accordingly. On the face of it, each one of these owes an account of what it is for something—an entity, a science—to be natural.

2.1 Ontological naturalism

Those entities an ontological naturalist is willing to accept will vary according to the conception of naturalness in play. On one view, an entity is natural if it is posited by those sciences typically considered natural—including, for example, physics, chemistry, biology, astronomy, and geology.[19] Some will find this characterization too broad, accepting only those entities posited by the most fundamental natural science, physics.[20] Others will find the characterization too narrow, and in addition accept those entities posited by the social sciences—perhaps including psychology, sociology, economics, and anthropology.[21] A quite different approach to ontological naturalism characterizes the natural, not by way of the (natural) sciences, but via some set of properties that all natural, and so acceptable, entities are claimed to possess. On one influential version of this view, natural entities are spatio-temporally located entities.[22]

[19] 'Ontological questions, under this view, are on a par with questions of natural science' (Quine 1948: 45).

[20] 'there is a further strand to my naturalism . . . physicalism, the thesis that all natural phenomena are, in a sense to be made precise, physical' (Papineau 1993: 1).

[21] Cf. Maddy's *Second Philosopher*: 'So, how *does* the Second Metaphysician proceed? For her the answer to "what is there?" takes the form of a list; what she actually confronts are a series of particular existence questions. What reason do we have, for example, to believe in the existence of medium-sized physical objects? . . . the answer is by now familiar: our chemical and physical story of such things as the apple on the table supports the commonsense view that the stuff making up the apple is importantly different from the stuff making up its surroundings; our electromagnetic and subatomic story explains how it holds together and resists penetration; our optics, physiology, cognitive science, biology and evolutionary theory describe how the underlying structures of our brain and nervous system react to light and other inputs from the apple to produce our belief in it. On these grounds, we hold that there is an apple there, that our belief that there is an apple is veridical' (Maddy 2007: 403–4).

[22] 'Naturalism I define as the view that nothing else exists except the single, spatio-temporal, world, the world studied by physics, chemistry, cosmology and so on' (Armstrong 1983: 82).

Obviously there is a great deal to be said about each of these varieties of ontological naturalism, and about the relations that hold between them. Spatio-temporal ontological naturalism immediately rules out both abstracta and supernatural concrete entities, such as God. Given that it seems unlikely that any of the natural or social sciences will posit the existence of supernatural entities, the other forms of naturalism will also disallow these. However, the issue regarding abstracta is not clear-cut. Whilst some have argued that our best scientific theories are committed to the existence of abstract objects (Quine 1954; Putnam 1975), others have denied this (Field 1980). Among those who would call themselves naturalists, both the best way to understand naturalized ontology and also what the particular ontological consequences of a naturalized ontology are, are live debates.

2.2 Methodological naturalism

Methodological naturalism takes a view about the appropriate methods for answering questions in any given domain, including those traditionally considered specifically philosophical. Those methods, the naturalist claims, are scientific methods.[23] That is, methodological naturalists reject the conception of philosophy as First Philosophy, a discipline that requires the natural and social sciences to answer to it. This orientation is well expressed by the Putnam of the early 1970s:

> It is silly to agree that a reason for believing that *p* warrants accepting *p* in all scientific circumstances, and then to add 'but even so it is not *good enough*'. Such a judgement could only be made if one accepted a trans-scientific method as superior to the scientific method: but this philosopher, at least, has no interest in doing *that*. (Putnam 1971: 356)[24]

The suggestion here is that a philosophical question, just like any other, ought to be answered by employing scientific techniques. As with ontological

[23] A weaker form of methodological naturalism maintains, not that the only legitimate methods are those of the natural sciences, but that in cases of conflict, natural scientific methods overrule any others. Cf. Burgess and Rosen (1997: 65): 'The naturalist's commitment is . . . to the comparatively modest proposition that when science speaks with a firm and unified voice, the philosopher is either obliged to accept its conclusions or to offer what are recognizably scientific reasons for resisting them.'

[24] Also, Quine's well-known remark, 'I hold that knowledge, mind and meaning are part of the same world that they have to do with, and that they are to be studied in the same empirical spirit that animates natural science. There is no place for a prior philosophy' (Quine 1968: 26).

naturalism, there is a question as to which sciences one means to include here. Are the techniques to be strictly those of the natural sciences—physics, chemistry, etc.—or are we to include the social sciences—psychology, sociology, etc.—as well? In the above quotation, Putnam seems to suggest that there is one, overarching method—the scientific method—which is to be employed in answering questions in all domains. If so, then which sciences are to be included will be answered by determining which sciences employ that method. However, there is a question mark over whether any such general way of demarcating scientific from non-scientific methods will be forthcoming.[25]

But methodological naturalism need not be held hostage to the demarcation problem in this way. On Maddy's view, a view she calls 'Second Philosophy', the correct method is just whichever technique—or techniques—works and stands up to critical scrutiny. As she puts it in the metaphysical case,

> The Second Philosopher conducts her metaphysical inquiry as she does every other inquiry, beginning with observation, experimentation, theory formation and testing, revising and refining as she goes, but without recourse to any official notion of what constitutes 'science', without any means of justification beyond her tried and true methods. (Maddy 2007: 411)[26]

Notice that whilst Maddy's list of methods is open-ended, it begins with observation. Indeed, it is plausible that any form of methodological naturalist will come into some conflict with strong claims as to the *a priori* status of philosophical knowledge, in particular if such knowledge is considered synthetic.

2.3 Epistemological naturalism

Epistemological naturalism is closely related to both ontological and methodological naturalism. The epistemological naturalist claims that knowledge is a natural phenomenon.[27] If ontological naturalism is true, all phenomena are natural phenomena. So epistemological naturalism is just a special case of that doctrine. If methodological naturalism is true, philosophical questions—including those of concern to the epistemologist—are to be answered using natural scientific methods. This, it might be

[25] See, for example, (Dupré 1993: Ch. 10).
[26] Also see Maddy's contribution to this volume.
[27] See (Kornblith 2002). Also see Kornblith's contribution to this volume.

supposed, gives the substance of the claim that knowledge is itself something natural. But there are a number of more specific ways of understanding what epistemological naturalism amounts to.

The first derives from Quine who, despairing of the prospects for what he saw as the traditional view of epistemology (effectively, foundationalism), proposed to replace traditional epistemology with psychology:

> Epistemology, or something like it, simply falls into place as a chapter of psychology and hence of natural science. It studies a natural phenomenon, viz. a physical human subject. (Quine 1969a: 82–3)

This is a radical view which has come in for a good deal of criticism.[28] It maintains that traditional epistemology ought to be abandoned in favour of an entirely empirical study of both how human subjects in fact arrive at the beliefs that they do and also the extent to which those beliefs are accurate. But less radical varieties of epistemological naturalism are also available. One more modest conception of epistemological naturalism holds that, whilst psychology does not *replace* traditional epistemology, the latter cannot proceed independently of the empirical study of the knowing subject. As a result, epistemological questions are not to be answered in an entirely *a priori* fashion.[29] As Kitcher has the naturalist ask,

> How could our psychological and biological capacities and limitations *fail* to be relevant to the study of human knowledge? How could our scientific understanding of ourselves . . . support the notion that answers to scepticism and organons of methodology . . . could be generated *a priori*? (Kitcher 1992: 58)

On this more modest picture, naturalized epistemology is the view that the theory of knowledge must contain some empirical element, plausibly drawn from psychology. However, it is not committed to the Quinean claim that epistemology is exhausted by the empirical study of cognition.

[28] For an influential argument that Quine's naturalized epistemology is not really epistemology at all, since it is non-normative, see (Kim 1988). Quine (1990: 19) denies that his naturalized epistemology is non-normative.

[29] It might be maintained that traditional epistemology inevitably makes some very general *a posteriori* assumptions. For example, non-naturalistic work in the epistemology of testimony might, without basing the claim on an *a priori* argument, make the assumption that more than one subject of experience exists. If this is right, then rejecting the view that all epistemological questions are to be answered entirely *a priori* will not yet qualify one as a naturalized epistemologist.

3. The natural and the transcendental

How, then, are transcendental philosophy and philosophical naturalism related? It might seem that, at least as far as transcendental idealism is concerned, the two are entirely at odds.[30]

Transcendental idealism seems committed, at least on some interpretations of it, to everything that the naturalist rejects. Transcendental idealism, at least on those interpretations, makes ontological claims. An obvious example concerns the existence of things in themselves:

> I grant by all means that there are bodies without us, that is, things which, though quite unknown to us as to what they are in themselves, we yet know by the representations which their influence on our sensibility procures us. (Kant 1783: 289)

This assertion, however, is not authorized by any natural or social science, and given Kant's claim that space and time are pure intuitions, is often interpreted as asserting the existence of a realm of non-spatial, non-temporal—and so in that sense non-natural—entities.

Furthermore, transcendental idealism makes essential use of a distinction between two levels of investigation and explanation: the empirical and the transcendental.[31] Empirical explanations—the sorts of explanations offered by the natural and social sciences—are perfectly in order for empirical purposes, but will not answer the questions posed by the transcendental philosopher—for example, how synthetic *a priori* judgements are possible (Kant 1781/7: B19). And it is the transcendental level of investigation and explanation that is considered fundamental. Certainly, at the empirical level, natural scientific methods can show us that space and time are real, and can determine their properties. However, at the more fundamental, transcendental level, we discover that space and time are ideal. This is exactly the sort of First Philosophical position that the methodological naturalist wishes to reject.

Finally, transcendental idealism seems entirely inhospitable to any variation on the project of naturalizing epistemology. Transcendental idealism is, at least in part, an epistemological doctrine. It makes claims about what we can and cannot know and, at its heart, is concerned to explain how certain sorts of knowledge are possible. It should be obvious that, at least in

[30] 'Whatever else is obscure about Kant's transcendental idealism, one thing is clear—it involves the rejection of naturalism' (Skorupski 1990: 7).

[31] For an instructive take on this distinction see (Bell 1999: 194–202).

Kant's case, none of this is carried out within the confines of a Quinean, or even a more moderate, form of naturalized epistemology. From a Kantian perspective, an attempt to draw the bounds of cognition via a reliance on the results of empirical psychology would seem entirely misguided. And far from thinking that epistemology must renounce its status as *a priori*, it is an explanation of the *a priori* that provides one of the main motivations towards transcendental idealism.

On the face of it, then, transcendental idealism is a profoundly non-naturalistic doctrine.[32] With respect to transcendental arguments the issues are less clear-cut. The most obvious respect in which a transcendental argument may appear naturalistically unacceptable concerns its deployment of premises that are (synthetic) *a priori*. Since many epistemological naturalists are sceptical of the possibility of *a priori* status (Kitcher 2000; Maddy 2000), the project of transcendental argumentation will immediately appear suspect. There might, however, be understandings of the *a priority* of transcendental claims that will lessen this initial tension.

Here it is useful to consider Kitcher's distinction between what he calls the 'official epistemological' conception and the 'tacit knowledge' conception (Kitcher 2006), each of which he claims to find in the first *Critique*. Roughly speaking, the official epistemological conception of the *a priori* is of an item of knowledge that is justified regardless of the particularities of one's experience (given, that is, that one has sufficient experience to acquire the relevant concepts). An example of the kind of knowledge that might satisfy this condition would be mathematical knowledge which, on Kant's view, is arrived at via a process of construction in pure intuition that is immune to being undermined by any possible experience.[33]

The tacit knowledge conception of the *a priori*, on the other hand, is that of a tacit belief, true at every world of which the subject can have

[32] To say this is not to say that transcendental idealism is anti-scientific. Kant himself was deeply engaged with the scientific achievements of his day (Friedman 1992), and one of the motivations towards transcendental idealism in the first *Critique* is the attempt to provide foundations for what Kant claimed were synthetic *a priori* principles within Newtonian physics (Kant 1781/7: B17–18).

[33] 'It must first be remarked that properly mathematical propositions are always *a priori* judgments and are never empirical, because they carry necessity with them, which cannot be derived from experience . . . I take first the number 7, and, as I take the fingers on my hand as an intuition for assistance with the concept of 5, to that image of mine I now add the units that I have previously taken together in order to constitute the number 5 one after another to the number 7, and thus see the number 12 arise' (Kant 1781/7: B15–16).

experience, that is not justified by experience but rather is necessarily employed in the making, and justification, of empirical judgements. As Kitcher puts it,

> The tacit knowledge that enables us to go beyond the bare materials given us through sensation to knowledge of the world that we actually experience is both causally crucial to the process of arriving at explicit judgments and epistemically required for them to be justified. (Kitcher 2006: 42)

It is controversial whether Kant relies on the tacit knowledge conception of the *a priori*, but an example might perhaps be the principle of causality.[34] This principle, tacitly held and holding true of every world of which we can have experience, would necessarily be brought to bear on experience and generate the possibility of the making of justified empirical judgements concerning cause and effect.[35]

If naturalists are sceptical of *a priority*, it is likely to be the official epistemological conception of *a priority*. None of the forms of naturalism mentioned above—ontological, methodological, or epistemological—need deny the existence of tacit knowledge operative in experience and the making of empirical judgement. Indeed, if the best cognitive science tells us that this well describes the workings of human cognition, then we ought to accept something along these lines.[36] In short, naturalists need not be empiricists in every sense of that term.

Of these two conceptions of the *a priori*, the official epistemological conception is far better suited to characterize the *a priority* of transcendental claims. For the tacit knowledge conception takes that which is *a priori* to be that which is tacitly believed and which is necessarily employed in the

[34] That mathematics and causation provide Kantian examples of the 'official epistemological' and 'tacit knowledge' conceptions of *a priority* respectively, is suggested by Kitcher: 'Typically, when he is concerned with mathematical knowledge, the official epistemological conception is paramount; when he is applying the transcendental method and analyzing the preconditions of cognition, as in the Analytic, the tacit knowledge conception comes to the fore' (Kitcher 2006: 52).

[35] 'it is only because we subject the sequence of the appearances and thus all alteration to the law of causality that experience itself, i.e., empirical cognition of them, is possible; consequently they themselves, as objects of experience, are possible only in accordance with this law' (Kant 1781/7: B234).

[36] 'from a thorough study of infant cognition and of evolutionary pressures, we might conclude that human brains come equipped to perceive medium-sized physical objects, and we might think it reasonable to describe ourselves as knowing something about the world . . . before experience, a priori. Of course, this is not at all what Kant has in mind!' (Maddy 2007: 63).

making of empirical judgements. But it is not plausible, nor is it any commitment of transcendental philosophy, that transcendental claims perform this role. Transcendental claims state the necessary conditions of experience. To suggest that these are *a priori* in the sense of Kitcher's tacit knowledge conception of the *a priori*, would be to suggest that each of us has beliefs *about* the necessary conditions of experience, beliefs that are operative in all empirical judgement. But this is far-fetched. It is the conditions themselves, not the belief that they are conditions, that are supposed to be operative in the making of empirical judgement. If transcendental claims are *a priori* in any sense, then, it seems clear that this will be the official epistemological sense, and that such claims will therefore remain problematic from certain naturalistic perspectives.

The correct response for such a naturalist will surely be to surrender altogether the *a priority* of transcendental claims, and to subsume them within an empirical and contingent theory of human cognition. But while this move has been a central step in many different twentieth-century attempts to reconcile transcendental philosophy with naturalism, the reconciliation is far from complete. We observed in §1 that transcendental philosophy is characterized by its explanatory aspirations, and for one who shares those aspirations the fact that naturalism must at the same time surrender the universality and necessity of transcendental claims serves only to highlight its explanatory deficits. It has been said in this vein, for instance, that the naturalist's methods can never yield complete or presuppositionless explanations (Bell 1999); that naturalists will inevitably have trouble accounting for normativity (Kim 1988); and that naturalists cannot successfully rebut the challenge of scepticism (Stroud 1984: Ch. 6).[37] On this latter point, the transcendental idealist will claim to have the upper hand. For, once again, part of the purpose of transcendental idealism is to answer certain epistemological questions—concerning the synthetic *a priori*, knowledge of the empirical world, etc.—that might appear naturalistically unanswerable.[38] In addition to the explanatory benefits of transcendental idealism, the transcendental philosopher might

[37] For a naturalistic response to Stroud's position, see (Maddy 2007: 20–36).

[38] Although, concerning the latter, there is the perennial concern, well expressed by McDowell, that, 'Once the supersensible is in the picture, its radical independence of our thinking tends to present itself as no more than the independence any genuine reality must have. The empirical world's claim to independence comes to seem fraudulent by comparison (McDowell 1996: 42). Also see Moore's contribution to this volume.

offer transcendental arguments against certain recognizably naturalist positions. Examples might be Stroud's (2004) argument against a 'restrictive naturalism' or Kitcher's argument, in her contribution to this volume, against certain naturalistic accounts of rational cognition.

Methodological and epistemological naturalists might contend in response that, if answering radical sceptical questions involves endorsing some form of transcendental idealism, then the price is too high; that the transcendental idealist's claim to have laid sceptical questions to rest is far from settled; and that the transcendental philosopher's use of *a priori* transcendental claims to refute naturalistic positions is question begging, at least to the extent that a naturalistic denial of the *a priori* has some independent motivation. As we noted, a conciliatory naturalist might then seek to co-opt some of the transcendental philosopher's results, by proposing naturalized transcendental arguments that employ premises drawn from the natural and social sciences.[39]

So there are various perspectives from which to view the relation between the transcendental and the natural. In very general terms we have seen that the transcendental philosopher is likely to resist both methodological and epistemological naturalism, presenting his or her own view as one that has fundamental explanatory advantages over the naturalistic orientation. The naturalist, on the other hand, will want to resist the separation of transcendental and empirical levels of explanation that is integral to the transcendental philosopher's project, perhaps by rejecting transcendental claims altogether, or perhaps by offering scientifically grounded reinterpretations and explanations of them. It is not the aim of this introduction to adjudicate between these perspectives, but only to provide a context for our contributors' more thorough and subtle explorations of them.

4. The chapters

In his contribution, Adrian Haddock discusses Davidson's (1974) argument against the possibility of alternative conceptual schemes and its

[39] There is a useful discussion of the prospects for naturalizing transcendental philosophy in (Cassam 2003). Also see Smith's contribution to this volume. The distinction between transcendental constraints and transcendental features, mentioned above, might also open up the possibility of varieties of transcendental idealism that lack some of the features offensive to various forms of naturalism.

relation to two versions of transcendental idealism: one Kantian; the other Wittgensteinian. According to Lear (1982, 1984) the later Wittgenstein maintains that, on the one hand, certain *a priori* knowledge—for example, knowledge of arithmetic—is grounded in 'how we are minded' and, on the other, that there is no alternative form of mindedness of which we can make sense. This latter contrasts with the Kantian view, according to which we cannot rule out the possibility of subjects with forms of intuition that differ from our own. Lear argues that Davidson is committed to a Wittgensteinian form of transcendental idealism. Without disputing this interpretation, Haddock argues that this would leave Davidson with an indefensible combination of views. Perhaps ironically, given Davidson's intentions, he would be better served by endorsing the claim distinctive of the Kantian variety of transcendental idealism.

In his contribution, A. W. Moore discusses Putnam's (1981) argument against the sceptical scenario that, for all I know, I am a brain in a vat.[40] Moore argues—by way of a contrast between a number of sceptical positions—that by adopting one particular way of resisting Putnam's argument 'to the last', one will end up with a position that amounts to a radical form of transcendental idealism. Transcendental idealism is, suggests Moore, an unattractive view. But it is perennially tempting, since we are drawn to the view that we have an inexpressible insight into the possibility that our thinking is only answerable to a limited aspect of the world, rather than to how things are in themselves—what Moore calls our 'phenomenal bubble'.

In her contribution, Patricia Kitcher presents an interpretation of the transcendental deduction as moving regressively from the assumption of empirical cognition to the necessary conditions of thinking, including the transcendental unity of apperception. Kitcher argues that the Kantian account of the unity of apperception shows the poverty of inner sense views. For Kant, only the conception of apperception as a type of 'act-consciousness' will guarantee that, in judging, the subject is aware of that very judgement as based on reasons. This unified awareness of judgement and reasons, argues Kant, is a necessary condition of a judgement's being *rational*. This, suggests Kitcher, has consequences for naturalistic attempts

[40] An argument which is itself often considered transcendental. See (Brueckner 1999).

to present computer models of the mind, for it is not clear how to incorporate Kantian act-consciousness into such a picture.

In his contribution, Robert Stern offers an interpretation and defence of Korsgaard's (1996) transcendental argument from our capacity for rational agency to the conclusion that we must value our own humanity. In Korsgaard's picture, to be capable of rational action one must take it that one has reasons to act and these must be considered reasons in the light of some practical identity that one adopts. This, Korsgaard argues, eventuates in the subject's being forced to recognize the value of their own humanity—the ultimate practical identity. Stern considers two ways of understanding Korsgaard's argument, finding both unconvincing. However, he argues that, on a third interpretation, the argument can be made compelling. That is, given the avowedly anti-realistic framework within which Korsgaard's argument is situated, one's valuing one's own humanity (valuing oneself qua rational agent) can be shown to be a necessary condition of the possibility of rational agency.

In his contribution, Hilary Kornblith contrasts two pictures of knowledge and epistemic reasons: one naturalistic; the other broadly Kantian. The naturalistic view, which Kornblith defends, takes knowledge and reasons as natural phenomena, to be investigated by the natural and social sciences. On the other hand, according to the broadly Kantian picture, such naturalization is seen to be unacceptable. In the Kantian picture humans, unlike other animals, have the capacity to reflect upon their own mental states and thus to take them as reasons for belief. In this way room is made for epistemic agency and we see the inadequacy of any naturalistic treatment of cognition as a kind of information processing. Kornblith presents a range of arguments against this broadly Kantian picture. He argues that the view is under-motivated, that it has empirical consequences that are demonstrably false, and that it makes it very hard to see how the normal course of human development could bring us from an animal-like situation into 'the space of reasons'.

In her contribution, Penelope Maddy articulates and defends the form of methodological naturalism that she calls 'Second Philosophy'. This view accepts the picture of the world and its contents generated by the natural and social sciences and remains sceptical of first-philosophical claims to the effect that such a picture is in some way inadequate to the tasks of philosophy—for example, the refutation of radical scepticism about the external world—and thus needs the validation of a non-natural, distinctively

philosophical form of enquiry. Maddy identifies Kantian transcendental philosophy as, perhaps, the paradigm case of such a first-philosophical position and—via a discussion of Kant's Refutation of Idealism—outlines her reasons for thinking that the move to such a two-level view is unmotivated. Maddy goes on to compare three varieties of therapeutic philosophy: that of Kant; that of Wittgenstein; and that of Austin. She argues that therapeutic philosophy—in particular the sort practised by Austin—can work alongside Second Philosophy in an attempt to justify putting questions, such as that concerning radical scepticism, to one side without answering them on their own terms.

In his contribution, Peter Sullivan offers a tentative positive answer to the question as to whether logic is transcendental. Sullivan argues that an 'easy' positive answer—by way of the thought that scepticism about logic is self-defeating—can appear unsatisfying, as it looks not to vindicate logic but to excuse it from vindication. In search of a better alternative, Sullivan offers a detailed reconstruction of a line of thought, and approach to the nature of transcendental arguments, to be found in (Sacks 2000). Broadly speaking, the strategy is to argue that certain structuring principles of thought are themselves the sources of answerability. Sullivan argues that it makes no sense to ask whether, in respect of its being structured by these principles, the mind is answerable to anything. It follows from this that logic is indeed transcendental. It also follows, suggests Sullivan, that given the rejection of a robustly realist setting according to which logic must be answerable to the way that things are in themselves, the initial 'easy' answer need not seem unsatisfying after all.

In his contribution, Joel Smith discusses Strawson's (1959) transcendental argument against scepticism about other minds. Smith defends the argument from a number of criticisms that have been levelled at transcendental arguments—criticisms that target the fact that transcendental arguments rely on synthetic *a priori* premises, that they purport to state facts about the world, and that they make claims of both necessity and universality. However, he goes on to argue that Strawson's argument does in fact fail since it confuses epistemological with developmental issues concerning our capacity to think about other minds. Smith then considers the possibility of a naturalized Strawsonian transcendental argument against scepticism about other minds. His conclusion, that Strawson's argument resists this naturalistic reinterpretation, illustrates the problems anticipated above in reconciling naturalist methods with the aims of transcendental philosophy.

References

Allison, H. 1969. Transcendental Idealism and Descriptive Metaphysics. *Kant-Studien* 60: 216–33.

—— 2004. *Kant's Transcendental Idealism: An Interpretation and Defence*, revised and enlarged edition. New Haven: Yale University Press.

Ameriks, K. 2003. *Interpreting Kant's Critiques*. Oxford: Clarendon Press.

Armstrong, D. M. 1983. *What is a Law of Nature?* Cambridge: Cambridge University Press.

Bell, D. 1999. Transcendental Arguments and Non-Naturalistic Anti-Realism. In Stern 1999: 189–210.

—— and Cooper, N. Eds. 1990. *The Analytic Tradition: Meaning, Thought, and Knowledge*. Oxford: Blackwell.

Boghossian, P. and Peacocke, C. Eds. 2000. *New Essays on the A Priori*. Oxford: Clarendon Press.

Brueckner, A. 1999. Transcendental Arguments from Content Externalism. In Stern 1999: 229–50.

Burgess, J. and Rosen, G. 1997. *A Subject With No Object: Strategies for Nominalist Interpretation of Mathematics*. Oxford: Oxford University Press.

Cassam, Q. 2003. Can Transcendental Epistemology be Naturalised? *Philosophy* 78: 181–203.

Davidson, D. 1974. On the Very Idea of a Conceptual Scheme. In Davidson 1984: 183–98.

—— 1984. *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

de Caro, M. and Macarthur, D. Eds. 2004. *Naturalism in Question*. Cambridge, MA: Harvard University Press.

Dupré, J. 1993. *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*. Cambridge, MA: Harvard University Press.

Field, H. 1980. *Science Without Numbers: A Defence of Nominalism*. Cambridge: Cambridge University Press.

Franks, P. 2005. *All or Nothing: Systematicity, Transcendental Arguments, and Skepticism in German Idealism*. Cambridge, MA: Harvard University Press.

Friedman, M. 1992. *Kant and the Exact Sciences*. Cambridge, MA: Harvard University Press.

Gardner, S. 1999. *Kant and the Critique of Pure Reason*. London: Routledge.

—— and Grist, M. Forthcoming. *The History of the Transcendental Turn*. Oxford: Oxford University Press.

Guyer, P. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

—— Ed. 2006. *The Cambridge Companion to Kant and Modern Philosophy*. Cambridge: Cambridge University Press.

Kant, I. 1781/7. *Critique of Pure Reason*. Translated by P. Guyer and A. Wood. Cambridge: Cambridge University Press, 1997.

—— 1783. *Prolegomena to Any Future Metaphysics That Will be Able to Come Forward as Science*. Translated by P. Carus and J. W. Ellington. Indianapolis: Hackett, 1977.

Kim, J. 1988. What is 'Naturalised Epistemology'? *Philosophical Perspectives* 2: 381–405.

Kitcher, P. 1992. The Naturalists Return. *Philosophical Review* 101: 53–114.

—— 2000. A Priori Knowledge Revisited. In Boghossian and Peacocke 2000: 65–91.

—— 2006. A Priori. In Guyer 2006: 28–60.

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

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Langton, R. 1998. *Kantian Humility: Our Ignorance of Things in Themselves*. Oxford: Clarendon Press.

Lear, J. 1982. Leaving the World Alone. *Journal of Philosophy* 79: 382–403.

—— 1984. The Disappearing 'We'. In Lear 1998: 282–300.

—— 1998. *Open Minded: Working Out the Logic of the Soul*. Cambridge, MA: Harvard University Press.

Maddy, P. 2000. Naturalism and the A Priori. In Boghossian and Peacocke 2000: 92–116.

—— 2007. *Second Philosophy: A Naturalistic Method*. Oxford: Oxford University Press.

McDowell, J. 1996. *Mind and World*, with a new introduction by the author. Cambridge, MA: Harvard University Press.

Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.

Peacocke, C. 2009. Objectivity. *Mind* 118: 739–69.

Pihlström, S. 2004. Recent Reinterpretations of the Transcendental. *Inquiry* 47: 289–314.

Putnam, H. 1971. Philosophy of Logic. In Putnam 1979: 323–57.

—— 1975. What is Mathematical Truth? In Putnam 1979: 60–78.

—— 1979. *Mathematics, Matter and Method: Philosophical Papers, Volume I*, 2nd edition. Cambridge: Cambridge University Press.

—— 1981. *Reason, Truth and History*. Cambridge: Cambridge University Press.

Quine, W. V. O. 1948. Two Dogmas of Empiricism. In Quine 1980: 20–46.

—— 1954. Carnap and Logical Truth. In Quine 1976: 107–32.

—— 1968. Ontological Relativity. In Quine 1969b: 26–68.

—— 1969a. Epistemology Naturalized. In Quine 1969b: 69–90.

—— 1969b. *Ontological Relativity and other essays*. New York: Columbia University Press.
—— 1976. *The Ways of Paradox*, revised and enlarged edition. Cambridge, MA: Harvard University Press.
—— 1980. *From a Logical Point of View*, 2nd edition. Cambridge, MA: Harvard University Press.
—— 1990. *Pursuit of Truth*. Cambridge, MA: Harvard University Press.
Sacks, M. 2000. *Objectivity and Insight*. Oxford: Clarendon Press.
—— 2005. The Nature of Transcendental Arguments. *International Journal of Philosophical Studies* 13: 439–60.
Skorupski, J. 1990. The Intelligibility of Scepticism. In Bell and Cooper 1990: 1–29.
Stern, R. Ed. 1999. *Transcendental Arguments: Problems and Prospects*. Oxford: Clarendon Press.
—— 2000. *Transcendental Arguments and Scepticism: Answering the Question of Justification*. Oxford: Clarendon Press.
Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge.
—— 1966. *The Bounds of Sense: An Essay on Kant's Critique of Pure Reason*. London: Methuen.
Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241–56.
—— 1984. *The Significance of Philosophical Scepticism*. Oxford: Oxford University Press.
—— 2004. The Charm of Naturalism. In de Caro and Macarthur 2004: 21–35.
Van Cleve, J. 1999. *Problems from Kant*. Oxford: Oxford University Press.
Walker, R. 2006. Kant and Transcendental Arguments. In Guyer 2006: 238–68.
Williams, B. 1974. Wittgenstein and Idealism. In Williams 1981: 144–63.
—— 1981. *Moral Luck: Philosophical Papers, 1973–1980*. Cambridge: Cambridge University Press.
Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London: Routledge, 1961.

2

Davidson and Idealism

Adrian Haddock

1. Donald Davidson (1974) argues against the possibility of conceptual schemes radically different from our own. Davidson intends thereby to argue against philosophical positions that presuppose this possibility, and he mentions Kant's transcendental idealism in this regard.[1] But there are those who think that his argument would establish idealism of some form, if it was sound. Thomas Nagel (1986: Ch. 6) thinks that if it was sound then it would establish a form of idealism which cuts reality down to the size of our concepts. Jonathan Lear (1982) thinks that if it was sound then it would establish transcendental idealism of a form associated not with Kant but with Wittgenstein.

Seeing that Nagel misunderstands Davidson is reasonably straightforward (§2). Seeing that there might be justice in Lear's thought is necessarily more circuitous. The form of idealism which Lear associates with Wittgenstein can be seen to make two central claims (§3). The first of these it shares with Kantian idealism; the second it does not (§4). Seeing that the soundness of Davidson's argument would establish the second of these Wittgensteinian claims is possible if we understand his argument in a certain plausible way (§§5–6). Unfortunately for Davidson, it seems that his argument is not sound (§7). Fortunately for Lear, it seems that the soundness of Davidson's argument would equally establish the first claim

[1] See (Davidson 1988).

central to Wittgensteinian idealism. Somewhat ironically, however, it seems that Davidson's argument would stand a better chance of establishing this first claim if Davidson espoused the second claim central to *Kantian* idealism (§8).

What follows, then, is a way of seeing how Lear's attribution of transcendental idealism to Davidson might be justified, as well as an argument that Davidson would do better to follow the way of Kant, rather than the way of Lear's Wittgenstein, if he is to realize some of the transcendental idealist ambitions which Lear takes him to harbour.

2. But let us begin with Nagel. Nagel endorses a certain form of realism:

> What there is and what we, in virtue of our nature, can think about are different things, and the latter may be smaller than the former . . . There are some things that we cannot now conceive but may yet come to understand; and there are probably still others that we lack the capacity to conceive not merely because we are at too early a stage of historical development, but because of the kind of beings we are. (Nagel 1986: 91–2)

Put differently, perhaps there are facts—propositions that are the case—which are unthinkable by us, 'because of the kind of beings we are'. Of course, we can think about these facts, in that we can think that there are facts which we cannot think. But we cannot think these facts.

Davidson's argument purports to establish that 'a form of activity that cannot be interpreted as language in our language is not speech behaviour' (Davidson 1974: 185–6)—is not *language* at all. And it purports to do this by establishing that the very idea of a form of activity which is language but is not interpretable as language in our language is an idea which makes no sense.

A form of activity interpretable as language in our language is a practice of expressing thoughts which admit of expression in our language. A form of activity interpretable as language but not in our language would be a practice of expressing thoughts which do not admit of expression in our language. Nagel assumes that by 'our language' Davidson just means a language which *we* are able to master in virtue of being 'the kind of beings we are'. He recognizes that Davidson places a prohibition on inexpressible thoughts, by identifying conceptual schemes with languages ('languages we will not think of as separable from souls' (Davidson 1974: 185)). So, he takes Davidson's argument to purport to establish that the very idea of thoughts which we are unable to think 'because of the kind of beings we are' is an idea which makes no sense. Nagel thinks the intelligibility of this

idea is a condition of the intelligibility of his form of realism. So, he thinks that Davidson's argument would (if sound) establish the unintelligibility of this form of realism. And he thinks that to establish *that* would be to establish a form of idealism.

Nagel's realism derives its plausibility from its account of 'the kind of beings we are'. We are human beings, members of a certain biological species, 'small and contingent pieces of the universe' (Nagel 1986: 92). I do not wish to dispute Nagel's assumption that Davidson shares this understanding of who we are.[2] But I do want to dispute his claim that the soundness of Davidson's argument would establish the unintelligibility of the idea of thoughts which we cannot think. Davidson's argument does purport to establish that the idea of a scheme no significant range of the sentences of which is translatable into or interpretable in our scheme is an idea which makes no sense. (Davidson uses the ideas of translation into and interpretation in interchangeably.) But Davidson does not purport to establish that the same is true of the idea of a scheme some range of the sentences of which is not translatable into our scheme. Exactly not; he wants to make sense of the idea of partial differences in scheme. To put it in Davidson's own terms: the idea of total failure of translation between our scheme and another is unintelligible; but the idea of partial failure of translation is not.

Davidson is famous for claiming that we can make sense of differences in belief only against a background of shared opinion. In the same way, he thinks we can make sense of differences in scheme only against a background of shared concepts: 'we must say much the same thing about differences in conceptual scheme as we say about differences in belief: we improve the clarity and bite of declarations of difference, whether of scheme or opinion, by enlarging the basis of shared (translatable) language or of shared opinion' (Davidson 1974: 197). That is not something he would say if he thought the idea of partial differences between our scheme

[2] Even though Davidson's reference to 'all mankind' (1974: 198) suggests that he shares this understanding, it is not decisive. But perhaps all Nagel needs is the thought that Davidson does not allow 'us' to shrink to a bare formal subject. And Davidson does not allow this. He thinks that his thesis—that a form of activity not interpretable as language in our language is not language—is not self-evident. But if 'we' shrank in this way, it would be self-evident (given his assumption that languages are essentially interpretable) because it would reduce to: a form of activity not interpretable as language is not language.

and other schemes was unintelligible; it is something he would say if he wanted to make sense of this idea.

Nagel's error is to assume that Davidson's argument targets the intelligibility of both the idea of total failure and the idea of partial failure. One final indication that only the intelligibility of the former is in his sights can be found in his description of his 'strategy [as] to argue that we cannot make sense of [the idea of] total failure, and then *to examine more briefly* cases of partial failure' (Davidson 1974: 185, my emphasis). He does not describe his strategy as to argue that we cannot make sense of the idea of partial failure.

3. Lear thinks that Davidson's argument establishes a 'Wittgensteinian form of transcendental idealism' (Lear 1982: 392, especially n.12), which Lear describes as follows:

> Let us say that a person is *minded* in a certain way, if he has the perceptions of salience, routes of interest, feelings of naturalness in following a rule, etc. that constitute being part of a certain form of life. And consider, for example, the alternative answers to the following question:
> What does 7+5 equal?
>
> (a) 12
> (b) Anything at all, just so long as everyone is so minded.
>
> To [this] question . . . (a) gives the correct answer. 7+5 equals 12, and anyone who tries to offer a different integer as an answer is in error. (Lear 1982: 385)

And yet, according to Lear, '[a]fter studying the later Wittgenstein, one is tempted to say that (b) also expresses some sort of truth' (Lear 1982: 386). But how can we believe what (b) says without that weakening our belief in the truth of what (a) says, and with it our belief in the necessity of addition?[3] The answer, according to Lear, and according to Wittgenstein, according to Lear, is by coming to see that there is no intelligible idea of being other minded.

The point may be put as follows. Transcendental idealism maintains that our *a priori* knowledge is grounded in the way we are minded. *A priori* knowledge is knowledge, not grounded in experience, of what is 'necessary and in the strictest sense universal' (B5). Our knowledge that 7+5 equals 12 is a paradigm of *a priori* knowledge. So, it is grounded in the way we are minded. To put it crudely: we are so minded that 7+5 equals 12, so

[3] See (Lear 1983: 45).

we know *a priori* that 7+5 equals 12. But this seems to entail that if there are thinking beings whose way of being minded is other than ours then these beings may enjoy *a priori* knowledge which differs from ours. To put it equally crudely: if there are thinking beings so minded that 7+5 equals 13, these beings will know *a priori* that 7+5 equals 13. That will threaten the status of our mathematical knowledge as knowledge of what is 'necessary and in the strictest sense universal'.

However, according to Lear, and according to Wittgenstein, according to Lear, this threat will vanish once we see that the idea of being other minded makes no sense—that the idea of being minded as we are is the idea of being minded *full stop*.

> Thus the strange case of the disappearing 'we'. [The] reflective understanding of the contribution of our mindedness to the necessity we find in the world is not meant to undermine the necessity, but to give us insight into it. (Lear 1984: 297)

The claim that our mindedness contributes to the necessity we find in the world would seem to undermine this necessity if, for all we know, there are other forms of mindedness. But Lear's thought is that if we know that the very idea of being other minded is unintelligible, the present threat to the status of our *a priori* knowledge as *a priori* knowledge dissolves.

So, the following two connected claims are central to the form of transcendental idealism which Lear associates with Wittgenstein: first, we can enjoy *a priori* knowledge grounded in how we are minded; second, we can see that there is no idea of being other minded.

4 . We can help to shed light on Wittgensteinian transcendental idealism by reminding ourselves of an integral argument of the Transcendental Aesthetic.[4]

The topic of the Aesthetic is the form of our sensible intuition. Intuition is 'that through which [cognition] relates immediately to [objects]' (A19/B33). This 'takes place only insofar as the object is given' (A19/B33) to the subject. And *this* is possible 'at least for us humans . . . only if [the object] affects the mind in a certain way' (B33). Intuition whose objects are given by means of affection—namely, 'by means of sensibility . . . the capacity (receptivity) to acquire representations through the way in which we are affected by objects' (A19/B33)—has a form which allows its objects 'to be

[4] Perhaps there is little hope of commanding widespread agreement in matters of Kant exegesis. I only hope that what I say is not unfamiliar. For comparable readings, see (Allison 2004) and (McDowell 2009).

ordered in certain relations' (A20/B34). And in the case of *our* sensible intuition, these relations are spatial and temporal.

Kant thinks that we enjoy *a priori* knowledge grounded in the form of our sensible intuition. He thinks we know *a priori*, on the basis of how our sensible intuition is formed, that all outer things are next to one another in space (for example). This makes it look as if thinking beings whose form of intuition is other than ours may enjoy *a priori* knowledge which differs from our own. To put it crudely: perhaps for them only some outer things are next to one another in space, just as for us only some outer things are coloured (for instance)—perhaps for them all outer things are in some other sort of relation. That will threaten the status of our *a priori* knowledge of outer things as *a priori* knowledge—as knowledge of what is 'necessary and in the strictest sense universal'.

Kant never doubts the intelligibility of the idea of differently formed intuitions. Indeed, he insists that 'we cannot judge at all whether the intuitions of other thinking beings are bound to the same conditions that limit our intuition and that are universally valid for us' (A27/B43). That is, we do not know whether there are thinking beings whose intuitions have forms which differ from the form of our intuition. So, we do not know—*inter alia*—whether all of the outer objects of the intuitions of other thinking beings are next to one another in space. We do not know whether there are thinking beings whose intuition is of outer objects only some of which (or indeed none of which) are in space. But if we do not know this how can we know *a priori*, on the basis of how *our* sensible intuition is formed, that *all* outer things are next to one another in space?

Kant has an answer to this question.

> The proposition: 'All [outer] things are next to one another in space,' is valid under the limitation that these things be taken as objects of our sensible intuition. If here I add the condition to the concept and say 'All things, as outer intuitions are next to one another in space,' then this rule is valid universally and without limitation. Our expositions accordingly teach the *reality* . . . of space in regard to everything that can come before us externally as an object [i.e., its empirical reality], but at the same time the *ideality* of space in regard to things when they are considered in themselves through reason, i.e., without taking account of the constitution of our sensibility [i.e., its transcendental ideality]. (A28/B44)

We can have *a priori* knowledge that all outer things are next to one another in space if these things are considered as appearances; as the objects of our sensible intuition. But we cannot know that all outer things are next to

one another in space if these things are considered as things in themselves; as things *simpliciter*, whether or not they are the objects of our sensible intuition.

So, the following two connected claims are central to Kantian transcendental idealism: first, we can enjoy *a priori* knowledge grounded in how we are minded (specifically, in the form of our sensible intuition); second, far from being able to see that there is no idea of being other minded, for all we know there are other forms of mindedness; consequently, our *a priori* knowledge must be restricted to appearances.

5. Lear introduces the form of idealism he associates with Wittgenstein via a mathematical example. But we can imagine a comparable treatment of the topic of outer spatiality, which insists that we enjoy *a priori* knowledge, grounded in our way of being minded, that all outer things are in space, which knowledge need not be restricted to appearances, because we can come to see that there is no idea of being other minded.[5] That provides a useful frame in which to understand Lear's claim that Davidson is a transcendental idealist of Wittgensteinian rather than Kantian stripe.

Lear sees Davidson's rejection of the dualism of scheme and content as central to Davidson's idealism. This dualism is capable of assuming a number of guises. The Kantian distinction between the categories (the pure concepts of the understanding) and the form of our sensible intuition is close to but not itself a guise of the dualism; the dualism comes with Kant's claim that the categories acquire content they would not otherwise possess when they are brought into contact with this form. The categories are not in themselves concepts of objects in space or time; but they become such concepts when they are applied to our sensible intuition. For example, whereas the category of cause and effect does not in itself contain the idea of temporality, our schematized category of cause and effect—the category in its relation to the form of our intuition—does contain this, for it is part of the schematized category that causes and effects are in time. This distinction between the categories in themselves and the categories as schematized by the form of our intuition is a guise assumed by the dualism in the *Critique*. To reject the dualism in this guise would be to identify the categories as such with (what figures in the dualism as) their schematized counterparts.

[5] This is not to say that mathematics and spatiality are unrelated, at least not in the *Critique*. The exegetical and philosophical difficulties which this putative relation raises are, to my mind, immense; see (Longuenesse 1998), especially Chapter 9.

It is not clear to me that Davidson does reject the dualism in this guise. However, I think a case can be made that he does reject one of its correlates, i.e., the intelligibility of the idea of differently schematized categories. I say 'correlate' and not 'consequence' because it is not clear to me that rejecting the intelligibility of this idea suffices to reject the dualism. Perhaps it is possible to distinguish between the pure content which the categories bear intrinsically, and that which they bear only in their relation to something outside of themselves, even if the spatial and temporal form of our sensible intuition is the only possible source of this pure but relational content. Nevertheless, because this idea is naturally seen as a guise of the idea of being other minded, if the soundness of Davidson's argument would suffice to establish that this idea is unintelligible, then there would be justice in attributing to Davidson the second of the two claims central to idealism of the form associated with Wittgenstein.

For Kant, it is part of the idea of a thinking being, as opposed to a being endowed with intellectual intuition, that thinking beings possess not merely the categories—of quantity, quality, cause and effect, and so on—but the categories schematized by a sensible form. But because Kant does not dispute the intelligibility of the idea of differently formed sensible intuitions, for him it is no part of the idea of a thinking being that thinking beings possess *our* schematized categories—the categories schematized by the spatial and temporal form of *our* sensible intuition. Our *a priori* knowledge might be thought of as grounded in our schematized categories, and thereby grounded in the form of our intuition. Our schematized category of cause and effect enables us to know that all causes and effects are in time. Our schematized category of an outer object enables us to know that all outer objects are in space. And so on. We have the first claim distinctive of transcendental idealism: our *a priori* knowledge is grounded in the way we are minded. Far from disputing the very idea of differently schematized categories, the Kantian idealist insists that for all we know there are differently schematized categories, because for all we know there are intuitions which are formed differently to our own. But he claims to dissolve the threat that this insistence seems to generate by restricting our *a priori* knowledge to appearances. A Wittgensteinian idealist, by contrast, would claim that we can see the meaninglessness of this idea. I think there might be justice in reading Davidson's argument as purporting to establish this Wittgensteinian claim.

6. Davidson's argument purports to establish that 'the *idea* of a radically foreign [conceptual] scheme' is 'meaningless' (Davidson 1988: 45). As we

have seen, he distinguishes between total and partial failure of translation, and argues against the intelligibility of the idea of the former. There 'would be [total] failure if no significant range of sentences in one language could be translated into the other; there would be partial failure if some range could be translated and another range could not' (Davidson 1974: 185). When Davidson speaks of a 'significant range of sentences' he might be taken to be speaking of a significantly large range of sentences. But the reading of Davidson, for which I think a case can be made, understands him as speaking of a range of significant sentences—sentences which express significant concepts, i.e., concepts with the status of Kantian categories.

Davidson sometimes says things which partially corroborate this reading. For instance he insists that, in denying the intelligibility of the idea of radically different schemes, he is saying that it is part of the idea of a language that languages have certain expressive powers, i.e., 'an underlying logical structure equivalent to the first-order predicate calculus with identity, an ontology of medium-sized objects with causal potentialities and a location in public space and time, ways of referring to the speaker and others, to places, to the past, to the present and future' (Davidson 1999: 308). This suggests that the significant sentences of a language are ones which express concepts associated with this range of powers: the concept of medium-sized objects; the concept of objects in space and time; the concept of objects with causal potentialities, and so on.

We can see Davidson here as taking issue with the Kantian assumption that it is no part of the idea of a thinking being that thinking beings possess our schematized categories. It is part of this idea that thinking beings possess the concept of medium-sized objects, if that is just the concept of objects of possible experience. But is it no part of the idea that this concept is so schematized as to be the concept of objects in space and time. And yet it can seem that for Davidson this is part of the idea, i.e., that it is part of the idea of a thinking being that thinking beings possess the category of objects of possible experience schematized by the form of our sensible intuition. The category of objects of possible experience is the most general category: every category is a category of objects of possible experience. So, it can seem that Davidson's point extends to all of the categories, i.e., that it is part of the idea of a thinking being that thinking beings possess the categories schematized by the form of our sensible intuition. And so, it can seem that Davidson denies the Kantian assumption.

That Davidson's argument does purport to establish that there is no intelligible idea of differently schematized categories is not a claim I wish

to endorse, but merely a way—perhaps the only way—of reading his argument which will allow us to see his argument as purporting to establish the second of the two claims distinctive of Wittgensteinian idealism, and thereby allow Lear's contention to stand a chance of being justified. I do not think this reading is entirely without foundation, but neither do I think it is clearly correct. Let us assume its correctness for the sake of argument, however, and ask whether Davidson's argument can establish this second claim.

7. Davidson argues that the intelligibility of the idea of a language radically different from our own requires 'a criterion of language-hood' (Davidson 1974: 186).

A criterion of language-hood is a way of telling—a way of knowing—whether a practice is a language rather than merely a patterned emission of noise (say). Davidson's claim that 'a form of activity that cannot be interpreted as language in our language is not speech behaviour' embodies just such a putative criterion. It is a putative criterion with two components: first, we know that if a significant range of the sentences of a practice are interpretable in our language, then the practice is a language; second, we know that if a significant range of the sentences of a practice are not interpretable in our language, then the practice is not a language.

This putative criterion rules out the possibility of a language which is not translatable in significant part into our language, because it entails that, if a significant range of the sentences of a practice are not interpretable in our language, then it is not a language. Davidson considers alternatives to this putative criterion, and argues that they should each be rejected. He concludes, first, that this putative criterion is the only viable criterion of language-hood, and, second—seemingly on the grounds that it is a condition of the intelligibility of the idea of a language which is not significantly translatable into our language, not just that there is a viable criterion of language-hood, but that this criterion does not rule out the possibility of such a language—that the idea of such a language is not intelligible.

We need not dwell on Davidson's reasons for thinking that there cannot be a viable criterion of language-hood which does not rule out this possibility in order to doubt whether an argument of this shape can be successful. First, even if there cannot be a viable criterion of language-hood which does not rule this out, this does not show that this putative criterion is the only viable criterion; perhaps there is no viable criterion of

language-hood. Second, it is hard to see how this putative criterion can be a viable criterion, given that what it entails is surely false. How can it follow from the fact that a significant range of the sentences of a putative language are not translatable into our language that this putative language is not a language? Surely the only thing that can follow is that *either* this putative language is not a language, *or* this putative language is a language which significantly resists translation into our own?

It might be objected that, if the upshot is that there is no viable criterion of language-hood, then Davidson's ultimate conclusion—i.e., the unintelligibility of the idea of a language which significantly resists translation into our language—is secured, because it is a condition of the intelligibility of the idea of a practice which significantly resists translation into our language but remains a genuine language that there is a way of telling whether a practice which significantly resists translation into our language is a genuine language. If there is no viable criterion of language-hood, then there is no such way of telling, and Davidson's conclusion is thereby secured.

But how can it follow from there being no way of telling whether an F is a G that the idea of an F's being a G makes no sense? Surely it can only be true that there is no way of knowing whether an F is a G if the idea of an F's being a G makes some sort of sense? It might be responded that to say that this idea makes no sense is merely to say that the concept of an F's being a G is like the concept of a married bachelor: such as to ensure that a claim to the effect that an F is a G entails a contradiction. But then why should it follow from the fact that there is no way of knowing whether a putative scheme radically different from ours is a genuine scheme that the claim that there is such a radically different scheme entails a contradiction? That inference is especially in need of defence in the present context, because it seems to violate Kant's dictum—central to his transcendental idealism—that '[t]o *think* of an object and to *cognize* an object are . . . not the same' (B146).[6] Concepts without intuitions are *empty*: they are not cognitions. To constitute cognition a concept requires an intuition, a way of being related to an object which enables one to get to know it. But concepts without intuitions are not *contradictory* ('my concept is a possible thought, even if I cannot give any assurance whether or not there is a

[6] Cognition is not the same as knowledge. But cognition and knowledge are connected, in that we can cognize something only if we can access an intuition which enables us to get to know it.

corresponding object somewhere within the sum total of all possibilities' (Bxxvi, n.)).

So, it is hard to see how Davidson's argument can enable us to see what the second claim essential to Wittgensteinian idealism maintains we can see, i.e., that there is no intelligible idea of being other minded. Of course, this does not undermine the claim that Davidson's argument would enable us to see this if it was sound. But it has a somewhat ironic consequence.

8. To see this consequence, we need to see that Davidson's argument does seem to purport to establish the *first* of the two claims associated with idealism in both its Kantian, and its Wittgensteinian forms, i.e., that we can enjoy *a priori* knowledge grounded in how we are minded.

The transcendental idealist can be rightfully represented as endorsing an inference from the fact that, as we know, all of the objects of our outer intuition are in space, to the claim that we know that all outer objects are in space. This inference is surely invalid. (That is one way to put the upshot of some of the considerations marshalled in §4 of this chapter.) Perhaps the most pressing problem is that, perhaps, for all we know, there are others who are differently minded—whose outer intuitions are differently formed. The second claim distinctive of the Wittgensteinian position attempts to supply the missing premise here: we know that there are no differently minded others, because we know that the very idea of being other minded makes no sense. The second claim distinctive of the Kantian position, by contrast, does not so much aim to supply the missing premise as to show us that we can have something like the upshot of taking the inferential step without actually doing so. We can *say* that, on the basis of how things are with our intuition, we know that every outer object is in space. But when we say that, what we mean is that we know that every outer object considered as an appearance is in space—and by that we mean every object of our outer intuition. Equivalently, what we mean is that every object of our outer intuition is in space. Even though our claim seems to go beyond the premise, and so beyond our intuition, to concern objects *simpliciter*, whether or not they are objects of our intuition, in fact it only concerns our intuition, because it restricts itself to our intuition, and to the premise. We might think of the Kantian position as saying that our *a priori* knowledge is a species of self-knowledge: it is knowledge, which is not derived from experience, of what is 'necessary and in the strictest sense universal'; but it concerns us, not things beyond us.

In the same way, Davidson can be rightfully represented as endorsing an inference from the fact that, as we know, every scheme which is significantly interpretable in our scheme embodies certain categories, to the claim that we know that every scheme embodies these categories. This inference seems to be undermined by the fact that, perhaps, for all we know, there are others who are differently minded: whose schemes embody different categories. Showing, and thereby enabling us to see, that the idea of such an alternative scheme is not intelligible would serve to supply the missing premise, in Wittgensteinian fashion. But it is interesting to note that a variant on the other, Kantian, response is also possible here. This would hold that, when we say that we know that every scheme embodies these categories, what we mean is that we know that every scheme considered as an appearance embodies these categories—where to consider a scheme as an appearance just means to consider it as a scheme which is significantly translatable into our own.

It is the fact that the transcendental idealist can be rightfully represented as endorsing the first of the two inferences outlined above which makes it right to ascribe the first claim to them, i.e., that we can enjoy *a priori* knowledge grounded in the way we are minded. That Davidson can be rightfully represented as endorsing a relevantly similar inference makes it right to do the same for him. In purporting to establish that every scheme embodies certain categories on the basis of the fact that *our* scheme embodies these categories—a fact which we know *a priori*—Davidson's argument purports to equip those of us who grasp his argument with *a priori* knowledge that every scheme embodies these very categories. Given our *a priori* knowledge of the former fact, the soundness of his argument would establish at least that we are in a position to acquire *a priori* knowledge grounded in how we are minded. (To move from being in this position to actual possession of the knowledge we simply need to grasp his argument.) Given this, and given what we saw in §§5 and 6 of this chapter, we can see that there might be justice in Lear's claim that Davidson is a transcendental idealist of the form associated not with Kant, but—by Lear as least—with Wittgenstein.

However, given what we saw in §7 of this chapter, it is hard to see how Davidson can justify the Wittgensteinian defence of the first idealist claim. Indeed, given what we have just seen, it might seem that he would do better to adopt the Kantian defence. At the very least, we are in the

position of seeing that Davidson might indeed endorse Wittgensteinian idealism, whilst being unable to see how he can do so defensibly.

But perhaps we can take this further. It seems that Davidson is left not only with a commitment to the second Wittgensteinian claim, which he fails to establish, but also with a commitment to the first claim, which he fails to establish precisely because he fails to establish the second claim. Short of either a new argument for this second claim, or an embrace—which according to Lear's reading he does not make—of the second claim distinctive of Kantian idealism, it is hard to see how he can establish the first claim. Thinking through what might be required for Lear's reading therefore leaves us in the position of saying that, if Davidson's commitment to the first claim is to be justified, it seems as if he must, contrary to the reading, embrace the second claim distinctive of Kantian idealism. We can have *a priori* knowledge that every scheme embodies our categories if schemes are considered as appearances, i.e., as schemes which are significantly interpretable in our scheme. But we cannot know that every scheme embodies these categories if schemes are considered as things in themselves, i.e., as schemes *simpliciter*, whether or not they are significantly interpretable in ours.[7]

This is not meant to be a criticism of Lear's reading. It does not suffice to impugn a reading to show that it attributes to its subject a claim that its subject cannot defend. My aim in this chapter has been to explore what is involved in reading Davidson's argument against radically different conceptual schemes as committing him to idealism. This exploration has led

[7] It *seems* as if Davidson must embrace the second Kantian claim. But perhaps there is an alternative. In his (2000), Mark Sacks argues that the apparent intelligibility of the idea of other-mindedness (in the shape of the idea of radically different 'normative structures') results from the combination of a form of naturalism—what he calls a 'post-metaphysical orientation'—with the assumption of an 'ontological base'. The base can only be mutable, given this orientation, and with the idea of its potential changes comes the idea of potentially different ways of being minded. This idea will be revealed as merely apparently intelligible if we can either reject this orientation, or reject this assumption. Rejecting the orientation is not an option for us moderns (or post-moderns). But, Sacks argues, we can and should reject the assumption. I do not dispute Sacks' diagnosis of why the idea of other-mindedness seems to be intelligible; but I cannot see how we can reject the assumption, in such a way as to remove this appearance of intelligibility, without also rejecting the orientation. The assumption of an ontological base seems to be in an important sense *part* of a post-metaphysical orientation—not something that can be simply removed from the scene, in such a way as to remove this appearance of intelligibility, whilst the orientation remains intact. There is a little more on Sacks' argument in my (2009). For a fuller treatment of Sacks (2000), see Peter Sullivan, 'Is Logic Transcendental?' (this volume).

me to reject the reading to this effect due to Nagel, but not the reading to this effect due to Lear.

Lear says that '[o]ne can . . . ironically, see Davidson as a type of transcendental idealist, even though he has done so much to oppose the scheme/content distinction' (Lear 1982: 392, n.12). It is a further irony, I think, that in the course of exploring whether Davidson's argument can succeed in establishing the idealist position which Lear ascribes to him, we have come to see that Davidson might have done better to embrace the form of transcendental idealism to which the dualism of scheme and content is essential—the form associated not with Wittgenstein, but with Kant.[8]

References

Allison, H. E. 2004. *Transcendental Idealism: An Interpretation and Defense*, 2nd edition. New Haven: Yale University Press.

Davidson, D. 1974. On the Very Idea of a Conceptual Scheme. *Proceedings and Addresses of the American Philosophical Association* 47: 5–20. Reprinted in his *Inquiries into Truth and Interpretation*: 183–98. Oxford: Clarendon Press, 2001. (Page references are to the reprinted version.)

—— 1988. The Myth of the Subjective. In M. Benedikt and R. Berger (eds), *Bewusstsein, Sprache und die Kunst*: 45–54. Vienna: Edition S. Verlag der ÖsterreichischenStaatsdruckerei. Reprinted in his *Subjective, Intersubjective, Objective*: 39–52. Oxford: Clarendon Press, 2001. (Page references are to the reprinted version.)

—— 1999. Reply to Simon J. Evnine. In L. E. Hahn (ed.), *The Philosophy of Donald Davidson*: 305–10. Chicago and LaSalle, IL: Open Court.

Haddock, A. 2009. McDowell, Transcendental Philosophy, and Naturalism. *Philosophical Topics* 37: 63–75.

Kant, I. 1997. *Critique of Pure Reason*. Translated and edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Lear, J. 1982. Leaving the World Alone. *Journal of Philosophy* 79: 382–403.

—— 1983. Ethics, Mathematics, and Relativism. *Mind* 92: 38–60.

[8] I would like to thank Peter Sullivan for his very helpful comments on an earlier draft. Thanks are also due to audiences at Cardiff, Liverpool, and Stirling for their very helpful comments, and to the Arts and Humanities Research Council for awarding me a period of research leave within which the integral work for this chapter was completed.

—— 1984. The Disappearing 'We'. *Proceedings of the Aristotelian Society Supplementary Volume* 58: 219–42. Reprinted in his *Open Minded: Working Out the Logic of the Soul*: 282–300. Cambridge, MA: Harvard University Press, 1998. (Page references are to the reprinted version.)

Longuenesse, B. 1998. *Kant and the Capacity to Judge: Sensibility and Discursivity in the Transcendental Analytic of the* Critique of Pure Reason. Translated by C. T. Wolfe. Princeton, NJ: Princeton University Press.

McDowell, J. 2009. Hegel's Idealism as Radicalization of Kant. In his *Having the World in View: Essays on Kant, Hegel, and Sellars*: 69–89. Cambridge, MA: Harvard University Press.

Nagel, T. 1986. *The View from Nowhere*. New York: Oxford University Press.

Sacks, M. 2000. *Objectivity and Insight*. Oxford: Clarendon Press.

3

Vats, Sets, and Tits*

A. W. Moore

1. At the beginning of Iris Murdoch's novel *The Black Prince* the narrator mentions four episodes from his story that might serve as suitable starting points and comments, 'There are indeed many places where I could start' (Murdoch 1975: 21). I feel a bit like that.

I could start with my title. Some explanation is certainly called for! I could start with a parenthetical remark that Hilary Putnam makes in *Reason, Truth and History*. Referring to his celebrated argument that we could not be brains in vats, Putnam says, 'This argument first occurred to me when I was thinking about a theorem in modern logic, the "Skolem-Löwenheim Theorem"' (Putnam 1981: 7). I could start with the very idea of critical self-conscious reflection on one's most basic beliefs about oneself and one's environment. I could start with my ultimate concern: transcendental idealism. Any of these would give my chapter a suitable steer—except, perhaps, the first, which would just allay curiosity. But the curiosity is liable to be distracting, so I shall in fact start there.

For as long as I can remember, the abbreviation that I have used in my note-taking for 'things in themselves' has been 'tits'. And, since this chapter will be concerned with comparisons and contrasts between three paradigms, in the third of which things in themselves play a role that is analogous to the role played by vats in the first and sets in the second,

* This chapter draws together ideas that I have expounded elsewhere, principally in (Moore 1996, 1997: Ch. 7, §1).

I found this title irresistible. But I note that the English word 'tits' has a number of meanings, even when not being used in what the dictionary would classify as a vulgar way. It can denote a kind of songbird of course; it can be used as a pejorative slang word to denote horses or young women; as a Scots word, it can denote twitches or tugs; and it can denote a gentle kind of knock, as in the phrase 'tit for tat'. In due course there will, I hope, be something agreeably apposite about this miscellany of very phenomenal associations.

2. To return to *Reason, Truth and History*. In Chapter 1 of that book Putnam considers the following scenario: a human brain is kept alive in a vat of nutrients and is manipulated by scientists in such a way as to give the subject the hallucinatory experience of living a perfectly normal life with a perfectly normal body. This scenario presents an obvious sceptical challenge: how do I know that I am not in this predicament? Putnam rises to the challenge by arguing that, at least in a sufficiently drastic version of the scenario,[1] its protagonist—let us call him Brain—cannot so much as entertain the thought that he is in this predicament, whatever other thoughts he might be able to entertain; which means that anyone who does entertain the thought that he is in this predicament thereby belies that very thought; which means that, when *I* entertain the thought that *I* am in this predicament, I belie my thought; which provides me with a way of meeting the sceptical challenge.

This argument has generated a large and fascinating literature.[2] One common objection to the argument is that, even if it gives me an assurance that I am *not* in the relevant predicament—an assurance, in other words, that my thought that I am just a brain in a vat (in sufficiently drastic circumstances) cannot be true—it does so only at the price of showing that I may not fully grasp this thought. *I may not know what this thought comes to at the level of things in themselves*. For, if Putnam's argument is correct, then Brain might entertain a thought *of the same type* as mine, with an assurance *of the same type* that his thought cannot be true, little realizing that, really, what his thought comes to is some very complex conjecture about a

[1] The reason for this qualification is that a good deal depends on what sort of causal contact there is, and has been in the past, between the brain and its environment: see (Smith 1984). Henceforth in this section I shall tend to take this qualification for granted, though in §3 below it will be crucial and I shall once again make it explicit.

[2] See e.g. (Brueckner 1986), (Sacks 1989: Ch. 3), (Wright 1994), to which Putnam replies in (Putnam 1994: §8), and (Forbes 1995).

possible configuration of things in a vat which, really, he is just a brain in. And the new sceptical challenge is: how do I know that I am not in a predicament analogous to *this*? As Crispin Wright puts it,

> [the] real spectre to be exorcized concerns the idea of a thought *standing behind* my thought that I am not a brain in a vat, in just the way that my thought that it *is* a mere brain in a vat would stand behind the thought . . . of an actual brain in a vat that 'I am not a brain in a vat' . . . What I should really like would be an assurance . . . not just that most of what I think is actually true . . . but that I am on to the right categories in terms of which to depict the most general features of the world and my place in it. (Wright 1994: 239–41)[3]

Anyone sympathetic to Putnam's argument is liable to say that this new sceptical challenge can be met in precisely the same way as the original one. Brain can no more entertain thoughts about what 'stands behind' his thoughts, or about 'the right categories' in terms of which to depict 'the most general features of the world', than he can entertain thoughts about brains and vats: that is, real brains and real vats.[4]

But those who see a genuinely new challenge here have two possible responses. The first of these is to distinguish between concepts such as that of a general feature of the world or that of a thing in itself, on the one hand, and concepts such as that of a brain or that of a vat, on the other, and to insist that Brain *can* entertain thoughts involving concepts of the former kind. This might be because concepts of the former kind are *a priori*, whereas concepts of the latter kind are empirical, so that possessing concepts of the former kind does not require the same sort of interaction with one's environment as possessing concepts of the latter kind does.[5]

[3] Adapted from the first-person plural to the first-person singular, his emphasis.

[4] Putnam himself responds to the new sceptical challenge in this way: see (Putnam 1994: 286–8).

[5] Cf. (Kant 1998: A85/B117). Note that, in order for this response to succeed, only a few concepts need to be of the former kind. And if they include the concept of analogy, then the response will not only succeed but succeed in a way that is very powerful, in the sense that it will significantly enlarge the range of sceptical thoughts that are available: consider, for instance, my thought that some being is to me in this or that respect as I am to Brain. See further (Moore 1996: §5), and the discussion there of (Nagel 1986). Cf. also my very use of the word 'analogous' in stating the new sceptical challenge above. For a very interesting discussion relevant to the question whether the concept of the self is of the former kind, and therefore possessable by Brain, see (Madden Forthcoming), where Madden raises some concerns about whether Brain can indeed think about, or refer to, himself. For remarks relevant to the possibility that he cannot, see (Evans 1982: 251–2). Also very relevant are Kant's remarks on self-consciousness at (Kant 1998: B157).

The second response is to grant Brain, and ultimately each of us, an insight into the possibility that there is some radical difference between how things are in themselves and how they appear which does not take the form of a normal thought at all; indeed, which might even be beyond our powers of expression.[6]

How different is this second response from the first? That depends on what counts as a 'normal' thought and on whether the insight in question *is* in fact supposed to be beyond our powers of expression. If a 'normal' thought is simply a thought that we can express, which is arguably just another way of saying that a 'normal' thought is simply a thought, and if, in accordance with that, the insight *is* supposed to be beyond our powers of expression, then the two responses are very different. If a 'normal' thought is a thought that involves concepts that are unavailable to Brain, because possessing them involves a certain sort of interaction with one's environment, and if the insight is *not* supposed to be beyond our powers of expression, then there may not be any difference at all. However that may be, both responses are gestures in the direction of a kind of transcendental idealism, whereby our 'normal' thoughts answer merely to how things appear, not to how they are in themselves. In Kant, from whom of course such transcendental idealism derives, the contrast between thoughts that are 'normal' in this way and thoughts that are not is the contrast between thoughts that have content and thoughts that do not, where this in turn is the contrast between thoughts that involve intuitions and thoughts that do not. Kant holds that thoughts of the former kind answer to how things are phenomenally: how they appear to beings with the relevant intuitions. These, for Kant, are the only thoughts that can constitute knowledge. But he allows for thoughts of the latter kind as well, abstract unverifiable conjectures about how things are in themselves.[7] And whatever we make of this contrast and its attendant metaphysics, it looks as

[6] Cf. (Moore 1997: Ch. 7, §1).

[7] See e.g. (Kant 1998: Bxxv–xxvii, 1996: Pt. I, Bk. II, Ch. 2, §VIII). John McDowell, in the opening section of (McDowell 1994), denies that Kant allows for thoughts of the latter kind: 'abnormal' thoughts, in the terms that I have been using. Commenting on the famous passage in which Kant declares that thoughts without content are empty (Kant 1998: A51/B75), McDowell writes, 'For a thought to be empty . . . would be for it not really to be a thought at all, and that is surely Kant's point; he is not, absurdly, drawing our attention to a special kind of thoughts, the empty ones' (pp. 3–4). But that is precisely what Kant is doing, or at least what he takes himself to be doing: see e.g. (Kant 1998: A253–4/B309).

though *that* is the sort of position that we shall end up in if we try to resist Putnam's argument to the last.[8]

3. To give these considerations somewhat sharper focus I shall now recast them in terms of three paradigms, which I shall call the Vat Paradigm, the Set Paradigm, and the Tit Paradigm.

The Vat Paradigm: I remarked above that Putnam's argument requires a 'sufficiently drastic' version of his scenario, but I did not elaborate.[9] Putnam himself envisages a universe in which *all* sentient creatures are brains in a vat being tended by automatic machinery that is programmed to give them a collective hallucination (Putnam 1981: 6). As it happens, less drastic versions of his scenario would have suited his purposes just as well. But there are versions that are less drastic still, for which his argument certainly fails. The Vat Paradigm is one of these. It concerns a human brain that has only *recently* been envatted and is being manipulated by scientists in such a way as to make the subject think that nothing untoward has happened. The subject in this case—let us call him Cerebrum—can certainly entertain the thought that he is in that predicament. If he does entertain this thought, and somehow reassures himself that it is false, then he is wrong.

The Set Paradigm: this concerns a set theorist—let us call him Georg—who uses standard set-theoretical terminology,[10] but under a non-standard interpretation. Georg's interpretation is elementarily equivalent to the standard interpretation,[11] in other words it makes precisely the same sentences come out true; but it differs from the standard interpretation in being a countable sub-interpretation of it. (It is the Skolem-Löwenheim theorem, the theorem to which Putnam refers in his parenthetical remark, that guarantees the existence of such an interpretation.) Georg is oblivious to the possibility of an interpretation extending his in the way in which the standard interpretation does. But, because his interpretation is elementarily equivalent to the standard interpretation, the sentences that he holds true, at least insofar as he is good at what he does, are the sentences that *are* true, under the standard interpretation. Among these is the sentence, 'There are uncountably many sets of finite ordinals.' There is a construction which, under Georg's

[8] To be sure, 'it looks as though' is the operative phrase. There are all sorts of twists and turns in the dialectic which I have ignored and which would have to be considered in any full discussion of these matters. Some philosophers think that Putnam's argument can be resisted at a much earlier stage: see e.g. (Nagel 1986: Ch. 5, §2). (Note that on p. 73 of Nagel's discussion, he considers a version of the second response to Putnam's extended argument.) For further reservations about how much Putnam's argument achieves see (Lewis 1984: 233–6).

[9] See above, footnote 1.

[10] The formal language in which he works, and in terms of which this terminology is defined, is the first-order language whose sole extralogical constant is ϵ.

[11] For current purposes I am simply taking for granted that there is such a thing as 'the standard interpretation'. Such an assumption is by no means philosophically innocuous, of course. But to address it would be a quite separate undertaking.

interpretation, constitutes a proof of this sentence, even though his interpretation has only countably many sets in its domain altogether.[12]

The Tit Paradigm: this concerns a subject—let us call him Noumenon—who views the world through native spectacles of some metaphorical kind which he can never take off. He has no knowledge of how things are in themselves, only of how they appear through the spectacles. But worse: he cannot even entertain or express thoughts about how things are in themselves. What he can do, however, is to achieve an insight, beyond his powers of expression, into the possibility that he is subject to precisely such limitations.[13]

These three paradigms have an important feature in common. In all three cases there is a subject whose thinking, in a certain respect, is sensitive only to a limited aspect of how things are in that respect: what I shall call the subject's 'phenomenal bubble'.[14] But there are also some important differences between the three paradigms. Cerebrum, unlike either Georg or Noumenon, is a victim of systemic error. Though his thinking is *sensitive* only to his phenomenal bubble, it is *answerable* to more than that. When he thinks, 'I am at the post office', he is thinking something that would be true only if he were at the post office, that is, only if things beyond his phenomenal bubble were a certain way. By contrast, Georg's thinking and Noumenon's thinking are each sensitive and answerable to the same thing: each is both sensitive and answerable to the thinker's phenomenal bubble. Neither Georg nor Noumenon need be involved in any error at all. In Georg's case, if not perhaps in Noumenon's, another subject can have thoughts *of the same type* as his that are answerable to how things are beyond his (Georg's) phenomenal bubble. For instance, I can think that there are uncountably many sets of finite ordinals. But although my set-theoretical thought is of the same type as one of Georg's, it does not have the same *content*.[15] That is why my thought and Georg's thought, despite being answerable to aspects of set-theoretical reality that differ in a

[12] For further details and discussion see (Moore 2001: Ch. 11).

[13] This is obviously meant to call to mind—I shall put it no more strongly than that—the metaphysical picture that Kant paints in (Kant 1998).(One reason for not putting it any more strongly than that is the point I made at the end of §2: Kant allows for thoughts about things in themselves.)

[14] For example, Georg's thinking, in respect of what sets are like, is sensitive only to what sets in the domain of his non-standard interpretation are like. Henceforth I shall normally take the qualification about the relevant 'respect' for granted.

[15] For amplification of this distinction between a thought's type and its content see (Moore 1997: 9–11).

way that might have been expected to confer different truth-values on them, are both in fact true. It is somewhat like the situation in which I watch a Wimbledon ball boy fumble a ball that has been gently lobbed to him, while Roger Federer watches his next opponent hit an unreturnable cross-court volley, and each of us thinks, quite correctly, 'I might have done that.' For that matter, it is somewhat like the situation in which I am at my desk, and Cerebrum is hallucinating being at his desk, and each of us thinks, quite correctly, 'I am *not* at the post office.' The difference is that, in the Set Paradigm, the relevant element of perspective is not located in anything as circumstantial as the use of the first person: it pervades the entire discourse. (And, of course, Georg, provided that he has arrived at his thought on sound mathematical grounds, is not *accidentally* correct in his thinking, in the way in which Cerebrum is.) Finally, the crucial difference between the Set Paradigm and the Tit Paradigm is that Noumenon, unlike Georg, can achieve an insight, albeit beyond his powers of expression, into the possibility that there is just such an element of perspective pervading everything he says and thinks; that everything he says and thinks is answerable only to his phenomenal bubble, not to things in themselves.

Corresponding to these three paradigms are three epistemological claims that we might make, which I shall refer to as vat-scepticism, set-scepticism, and tit-scepticism. Vat-scepticism is the claim that, for all I know, I am in Cerebrum's situation; set-scepticism is the claim that, for all I know, I am in Georg's situation; and tit-scepticism is the claim that, for all I know, I am in Noumenon's situation.

Now it is natural to assimilate ordinary philosophical scepticism concerning the truth of one's most basic beliefs about oneself and one's environment to vat-scepticism. It is natural, in other words, to think that the target of such scepticism is the possibility that one is a victim of systemic error of the sort that afflicts Cerebrum. And Putnam's argument, as we have noted, does nothing to rebut vat-scepticism. What it does do, however, if successful, is to call into question the assimilation, by showing how limited vat-scepticism is. In philosophical terms, Cerebrum's situation is not particularly outlandish. One would not have to have an especially philosophical cast of mind to think that vat-scepticism could *not* be rebutted; indeed, that it was true. But if Putnam is right, the more drastic possibilities of concern to philosophical sceptics are closer, in various critical respects, to the Set Paradigm than they are to the Vat Paradigm: in particular, Brain's predicament is closer, in these respects, to Georg's

predicament than to Cerebrum's. (Hence Putnam's remark about the Skolem-Löwenheim Theorem.)

To put it in these terms, however, just seems to reinforce the original objection to Putnam's argument, which was that it answers one sceptical challenge only by presenting a new one. The new challenge can now be seen as a variation on *set*-scepticism. For suppose that Putnam is right and I am not a victim of systemic error of any drastic kind in my thinking about the external world. How do I know that this is not just because my thinking is answerable to nothing but my phenomenal bubble, in a way of which I have not the least idea? Call this variant of set-scepticism set-like-scepticism.

Now set-scepticism itself—never mind, for the time being, any such non-mathematical variant of it—can certainly be rebutted. This is something that I have argued elsewhere, as indeed has Putnam, to whom I am indebted.[16] The argument, in summary, runs as follows.

The Argument Against Set-Scepticism: Set-scepticism is the claim that, for all I know, I am in Georg's situation, which entails that, for all I know, my thinking about sets is never thinking about *all* sets, but only ever about those in the domain of my own restricted interpretation of the language of set theory. But even in entertaining the thought that I am in that situation, I belie it. For it is itself a thought about *all* sets.

The set-sceptic will of course caution that these references of mine to 'all' sets may themselves have limited scope. But what does this mean? That they may not really be references to *all* sets? Yet that is precisely what they are!

Is there an analogue of this argument serving to rebut set-like-scepticism? Apparently so. It runs as follows.

The Argument Against Set-Like-Scepticism: set-like-scepticism entails that, for all I know, my thinking about the external world is never answerable to the external world in *all* its aspects, but only ever to my phenomenal bubble. But even in entertaining the thought that I am in that situation, I belie it. For it is itself a thought that is answerable to the external world in *all* its aspects.

Does this argument really serve to rebut set-like-scepticism? It may very well do, but arguably only at the price of showing that set-like-scepticism is not the scepticism that really concerns us; or in other words, that the assimilation of Brain's predicament to Georg's needs to be questioned. For Brain's predicament, however much like Georg's it may be, and in

[16] (Moore 2001: Ch. 11, esp. §3) and (Putnam 1983).

particular however much more like Georg's it may be in certain critical respects than it is like Cerebrum's, is nevertheless different from Georg's in other respects that are just as critical.

To amplify. The argument against set-like-scepticism is very similar to, if not identical with, the extension of Putnam's argument that we considered in the previous section. But when we considered that extension, we also considered two responses to it. And precisely what these responses both did, in effect, was to call into the question the assimilation of Brain's predicament to Georg's. The first response suggested that they are unlike each other in the following respect: not *all* of Brain's thoughts are answerable only to his phenomenal bubble. The second response, at least in its more radical forms, conceded that Brain's predicament is like Georg's in that respect—Brain's thoughts are indeed all answerable only to his phenomenal bubble—but fastened on a different difference. It suggested another shift of paradigm in fact: from the Set Paradigm to the Tit Paradigm. For the basic idea was that, even if I cannot *think* that my thoughts are all answerable only to my phenomenal bubble, still I can achieve an insight of sorts, albeit beyond my powers of expression, into the possibility that they are.[17] And if I *am* able to achieve such an insight, then we seem left with a final, unanswered sceptical challenge: a variation, this time, on *tit*-scepticism.

Or is 'sceptical challenge' the right description? 'Tit-scepticism' itself is just a label. It is a substantive question whether the claim to which that label attaches (namely, that, for all I know, I am in Noumenon's situation) deserves to count as a species of *scepticism* at all. Might it not be better viewed as a gesture towards a familiar and redoubtable philosophical doctrine: transcendental idealism? Which of course brings us back to the position that we were in at the end of the previous section.

4. I am in no doubt that we have just witnessed one of the main impulses towards a radical version of transcendental idealism. The idea that I have some kind of insight into the impossibility of my thinking

[17] It is worth noting parenthetically that these are by no means the only reasons for resisting the assimilation of Brain's predicament to Georg's. Someone might argue, for instance, that there is nothing more to sets than what we think about them, in contrast to the things that Brain thinks about. Again, relatedly, someone might argue that there is nothing more to how a set theorist interprets the language of set theory than which sentences from that language he or she holds true, which means that the Set Paradigm is incoherent, or in other words that Georg's predicament is not a genuine possibility. But not even Putnam would say this about Brain's predicament: see (Putnam 1981: 7–8).

about things in themselves, and into the corresponding limitations of that which I *can* think about, is a very natural destination for the train of thought that departs from Putnam's own original argument.[18]

Is it an attractive destination? My own view is that it is not. If we are to board this train at all—and I think there are good reasons for doing so—then I believe we should alight at an earlier stage, earlier even than the stage of conceding that, *for all I know*, such transcendental idealism is true. Come to that, I believe we should alight before acceding to a less radical version of transcendental idealism whereby the insight that I have into my own limitations is one that I can express, and the limitations are limitations only to that part of my thinking which is 'normal' in some robust sense.[19] I believe we should alight at the argument against set-like-scepticism.[20]

I cannot argue for this now. To do so would require saying considerably more than I have been able to say in this chapter about why we should go even that far. But I shall close by saying a little about why I do not think that we should go all the way to the more radical version of transcendental idealism, even though there is a strong temptation to do so.

Part of the reason why there is a strong temptation to do so, it seems to me, is that we do have an inexpressible insight, activated by Putnam's argument, into what it is for that to which our thinking is answerable to be how we think it is; and, if we try to express this insight, then we are liable to circumscribe that to which our thinking is answerable and treat it as some kind of phenomenal bubble. We are liable to endorse the more radical version of transcendental idealism. But while this explains our temptation, it does not in any way vindicate it. The fact is, we cannot treat that to which our thinking is answerable as some kind of phenomenal bubble, for a reason famously articulated by Wittgenstein: 'in order to be able to draw a limit to thought, we should have to find both sides of the limit thinkable (i.e., we should have to be able to think what cannot be thought)' (Wittgenstein 1961: 3). (This is reminiscent of the argument

[18] Putnam himself is not averse to the idea that there is a route from his argument to a kind of transcendental idealism: see e.g. (Putnam 1981: 60–64). Cf. also (Sacks 1989: Ch. 3) though Sacks construes transcendental idealism in a more epistemological way than I have been doing—as something very like tit-scepticism in fact.

[19] See again the material at the end of §2 above: Kant's own transcendental idealism is of this less radical kind.

[20] This means, in particular, that we should resist the attempts above to dissociate Brain's predicament from Georg's.

against set-like-scepticism, the point at which I have already said I think we should be stationing ourselves: I cannot entertain the thought that my thinking is answerable only to my phenomenal bubble without having a thought that is answerable to more than my phenomenal bubble.)

But still, the transcendental idealist may say, if we really do *have* the inexpressible insight, and thereby share with Noumenon the crucial feature that distinguishes him from Georg, what is to stop us from simply alluding to that fact and thus assimilating our own case to the Tit Paradigm? And if we do that, shall we not be *en route* to the more radical version of transcendental idealism after all?

Well, unless we actually try to express the insight, not just allude to the fact that we have it, we shall *not* have assimilated our own case to the Tit Paradigm. To share with Noumenon the crucial feature that distinguishes him from Georg, it is not enough that we have an inexpressible insight; obviously not. It is not even enough that we have an inexpressible insight that we are tempted to treat as an insight into the possibility that our thinking is answerable only to our phenomenal bubble. We need to have an insight that *really is* an insight into that possibility. And—quite apart from the fact that any insight that really was an insight into that possibility would *eo ipso* be expressible—if there is no such possibility, then there can be no such insight. So the most we can do is to acknowledge our own temptation to endorse the more radical version of transcendental idealism. And that clearly falls short of actually endorsing it. It even falls short of conceding its coherence.

If I am right that we should not endorse any such transcendental idealism, and if I am right about how far short of that we should allow this train of thought to take us, what does this mean, finally, about things in themselves? It certainly removes one substantial reservation that we might have about whether we can so much as think about things in themselves. But we should not conclude without further ado, as we might be tempted to, that things in themselves are just ordinary middle-sized dry goods and their like. It is possible, for instance, that the phrase 'things in themselves' is best viewed as syncategorematic, so that 'thinking about things in themselves' means something like 'thinking about things in a way that is totally free of perspective, whether cultural, biological, or of any other kind', in which case we may have to conclude that things in themselves 'are'—to the extent that it is appropriate even to talk in these

terms—the fundamental particles of physics or something of that sort.[21] Still, even if things in themselves should not be thought of as including ordinary middle-sized dry goods, or the various episodes that involve them, we have some kind of reassurance that they can legitimately be regarded as being *of a piece with* ordinary middle-sized dry goods and the various episodes that involve them—such as songbirds, horses, and gentle knocks, things of the sort that can be denoted by the English word 'tits'.

References

Benacerraf, P. and Putnam, H. Eds. 1983. *Philosophy of Mathematics: Selected Readings*. Cambridge: Cambridge University Press.

Brueckner, A. 1986. Brains in a Vat. *Journal of Philosophy* 83: 148–67.

Clark, P. and Hale, B. Eds. 1994. *Reading Putnam*. Oxford: Blackwell.

Evans, G. 1982. *The Varieties of Reference*. Edited by J. McDowell. Oxford: Oxford University Press.

Forbes, G. 1995. Realism and Skepticism: Brains in a Vat Revisited. *Journal of Philosophy* 92: 205–22.

Kant, I. 1996. *Critique of Practical Reason*. In Kant, *Practical Philosophy*. Translated and edited by M. J. Gregor. Cambridge: Cambridge University Press.

—— 1998. *Critique of Pure Reason*. Translated and edited by P. Guyer and A. W. Wood. Cambridge: Cambridge University Press.

Lewis, D. 1984. Putnam's Paradox. *Australasian Journal of Philosophy* 62: 221–36.

Madden, R. Forthcoming. Could a Brain in a Vat Self Refer? *European Journal of Philosophy*.

McDowell, J. 1994. *Mind and World*. Cambridge, MA: Harvard University Press.

Moore, A. W. 1996. Solipsism and Subjectivity. *European Journal of Philosophy* 4: 220–34.

—— 1997. *Points of View*. Oxford: Oxford University Press.

—— 2001. *The Infinite*, 2nd edition. London: Routledge.

Murdoch, I. 1975. *The Black Prince*. Harmondsworth: Penguin.

Nagel, T. 1986. *The View From Nowhere*. Oxford: Oxford University Press.

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

—— 1983. Models and Reality. In Benacerraf and Putnam 1983: 421–44.

—— 1994. Comments and Replies. In Clark and Hale 1994: 242–95.

Sacks, M. 1989. *The World We Found: The Limits of Ontological Talk*. London: Duckworth.

[21] Cf. (Moore 1997: Ch. 4, §4).

Smith, P. 1984. Could We Be Brains in a Vat? *Canadian Journal of Philosophy* 14: 115–23.

Wittgenstein, L. 1961. *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London: Routledge.

Wright, C. 1994. On Putnam's Proof that We Are Not Brains in a Vat. In Clark and Hale 1994: 216–41.

4

The Unity of Kant's Active Thinker

Patricia Kitcher

Introduction

Kant thought that he could establish conclusions about the mind and nature by making transcendental arguments; he believed that his results had anti-empiricist, if not anti-naturalistic, implications about the mind. The themes of transcendental philosophy, naturalism, and the mind are closely tied together in his work. This is no accident. It is largely because of our Kantian heritage that we think of the three issues as interconnected. Over the last forty years or so, many philosophers have believed that a Kantian approach could be employed in the service of systematic philosophy. I have three aims in the chapter. First, I try to clarify the Kantian legacy by considering more closely how these issues were connected in his work. Second, I draw on that analysis to show that Kant's transcendental arguments had at least one anti-mechanistic and two anti-empiricist implications about the mind. Third, I argue that he achieved these results because he made stronger assumptions about the psychological prerequisites of cognition than most contemporary philosophers are willing to countenance. If I am correct, then some of Kant's substantive theses may be as useful to contemporary epistemology and philosophy of mind as his distinctive transcendental method.

Section 1 gives a brief overview of how the transcendental deduction (hereafter 'TD') is supposed to work. The second and longest section presents an account of a central piece of the TD, the argument for transcendental apperception. I discuss this argument in some detail to

show that it involves features that are not shared by other transcendental arguments. In section 3 I bring out the strength and uniqueness of Kant's argument for the unity of transcendental apperception by contrasting it with a very sophisticated contemporary attempt to show that object cognition requires a continuing thinker—that offered by Quassim Cassam. In the final section, I argue that in exploring the requirements of cognition, Kant discovers a kind of consciousness that both plays a crucial role in mental unity and is not obviously explicable in terms of current naturalistic models of mind. In David Chalmers' (1996) terminology, it is a kind of consciousness that raises a 'hard problem' for contemporary theories, even though it is unrelated to standard cases of the 'hard problem', cases such as the unpleasant quality of pains or the ineffable visual quality of a purple haze. The last two sections thus argue that Kant's TD has implications for contemporary work on the mind, specifically for theories of mental unity and for naturalistic theories of consciousness.

1. What is a transcendental deduction?[1]

Transcendental arguments have interested recent philosophy because of their seeming potential to defang various sorts of scepticism. Two famous types of scepticism thought vulnerable to Kantian-type arguments were scepticism about external objects and scepticism about other minds. On the other hand, a number of historians of philosophy beginning (I believe) with Margaret Wilson (1974) have denied that Kant had Cartesian scepticism in his sights. His target was not Descartes, but Hume. Kant's description of the problem for *The Critique of Pure Reason* supports the historians. His project was to explain

How are synthetic *a priori* judgments possible? (A19, cf. A9)[2]

He defines *a priori* cognition in contrast to the *a posteriori* variety in the Introduction to the First Critique.

[1] Some of the material in this section appeared in (Kitcher 2008).

[2] References to the *Critique of Pure Reason* will be in the text, with the usual 'A' and 'B' indications of editions. In providing English translations, I usually rely on (Pluhar 1996), but I also borrow freely from (Kemp Smith 1968), and from (Guyer and Wood 1998), and sometimes I combine translational suggestions from the different standard references. Where I substantially alter a translation, I indicate that the translation is amended. I also use Guyer and Wood's convention of indicating Kant's emphasis through boldface type. References to Kant's works, other than the first *Critique*, will be to (Kant 1900–) and will be cited in the text by giving volume and page numbers from that edition.

But even though all our cognition commences **with** experience, nevertheless, it does not for that reason all originate **from** experience. For it might well be that our empirical cognition itself is a composite of what we receive through impressions and of what our own cognitive faculties give up out of themselves (merely induced by sensory impressions). (B1)

One calls such **cognitions** [i.e., what our cognitive faculties give up out of themselves] **a priori,** and distinguishes them from the **empirical** [ones], which have their sources *a posteriori*, namely in experience. (B2)

Since the *a priori* elements of cognitions come not from objects, but from activities of the mind, there is a special problem in establishing the legitimacy of their use.

Kant believed that in order to vindicate the use of *a priori* concepts he needed to develop a new type of argument, a TD. He was explicit about the unique feature of such a deduction:

the transcendental deduction of all *a priori* concepts has a principle to which the entire investigation must be directed: *viz*., the principle that these concepts must be recognized as *a priori* conditions for the possibility of experience (whether the possibility of the intuition found in experience, or the possibility of the thought). (A94/B126)

In transcendental cognition, so long as we are concerned only with the concepts of the understanding, our guide is the possibility of experience . . . The [transcendental] proof proceeds by showing that experience itself, and therefore the object of experience, would be impossible without a connection of this kind [between concepts]. (A783/B811)

These descriptions immediately invite the question: What is the 'possibility of experience'?

On this point, Kant's published and unpublished remarks are clear. The possibility of 'experience' should be understood as the possibility of 'empirical cognition':

The categories serve only for the possibility of **empirical cognition**. Such cognition, however, is called **experience**. (B147, see also, e.g., 7.141, 18.318)

And by the 'possibility of empirical cognition', he means the possibility of attaining cognition of objects through receiving information through the senses.[3]

Kant signals that the assumption that cognition through the senses is possible for creatures like us is an acceptable starting place for his defence of

[3] Carl (1989, 1992) argues against this view. See (Kitcher 2011: Ch. 7) for discussion.

the categories in the opening sentence of the Introduction (the sentence just preceding the material cited above):

> There can be no doubt that all our cognition begins with experience [with the senses rousing our cognitive power to its operation]. (B1)

The assumption of the possibility of empirical cognition is a uniquely appropriate starting place for a defence of *a priori* cognition. He would be able to argue that *a posteriori* cognition, against which the *a priori* is unfavourably compared, itself requires *a priori* contributions from the mind.

We seem to have a simple account of the argument. It starts from the assumption of empirical cognition and then regresses to certain *a priori* concepts whose use is shown to be a necessary condition for the possibility of empirical cognition. This is too simple in some respects, but I will add only two complications. As Kuehn (1997) among others notes, at a crucial point in the set-up to the TD, Kant explained his project in terms of 'objective validity':

> With the categories of the understanding we encounter a difficulty that we did not encounter in the realm of sensibility: viz. how **subjective conditions of thought** could have **objective validity**. (A89/B122)

Kuehn connects Kant's description of his problem to the theories of Christian Wolff (1751/1983). Wolff understood philosophy as 'the science of all possible objects, how and why they are possible' (Kuehn, 1997: 229). On Kuehn's account, Wolff's goal was to prove that certain concepts were possible, by showing that objects falling under the concepts were possible (1997: 232). Kuehn notes that G. F. Meier's logic text, which Kant used as a basis for his lectures, laid out the basics of the Wolffian programme:

> A learned concept which has been created by arbitrary conjunction must be proved or disproved. We can achieve this either by (i) experience, if we show that their concepts are real or not real, or (ii) by reason, either directly or indirectly by showing that and how their objects can become real or that they cannot become real. (cited in Kuehn 1997: 235)

Suppose that Bart has the concept 'ooblick', the concept of green sticky stuff that falls from the sky like rain. He could prove the possibility of his concept either by seeing some ooblick (not likely) or by explaining how ooblick might come into existence. In light of the Wolffian background, we can understand why Kant had no worries about empirical concepts.

Their legitimacy or objective validity could always be established whenever a doubt arose (A84/B116–17). We can also see why sensitive readers such as Strawson thought that the success of the TD must imply not just the necessity of using object concepts, but the existence of 'objects' in the sense captured by the *a priori* concept of something that has properties and undergoes change through its interaction with other things.

The second complication comes from Kant's famous analogy between the TD and legal deductions. Oddly, this clue remained unexplored until Henrich's (1989) pioneering study. As he explains, the practice of deduction-writing arose because some means were needed to settle disputes about property and inheritance among the various entities that had constituted the Holy Roman Empire. He explained how they worked:

> In order to determine whether an acquired right was real or only presumption, one must legally trace the possession somebody claims back to its origin. The process through which a possession or a usage is accounted for by explaining its origin, such that the rightfulness of the possession or the usage becomes apparent, define the deduction. (Henrich 1989: 35)

As he notes, the analogy with legal deductions explains the TD's numerous references to the 'origins' of representations. In the preparatory section to the TD, Kant characterized one of its aims as providing for the categories 'a *certificate of birth* quite other than descent from experience' (A86/B199, my emphasis).

Kant assumed that there were only two possible origins for representations. Either they arose through outer causes operating through the senses or through inner causes, that is, through the activities of the mind (A98). The only alternative to an empirical birth certificate would be to trace the origin of the categories back to the operations of mental faculties. In particular, Kant tries to show that categorial concepts arise from activities of the mind that are necessary for any thinking at all. That is, on analogy with legal deductions, the TD traces the usage of certain concepts back to operations of the mind that are necessary in all thinking, thereby providing an origin for the concepts that is suitable to their role, which is that they are applicable to all objects of cognition. An empirical derivation could show that the concept in question applies to some objects, but a TD is supposed to show why the categories apply to any object that can be thought at all. This analysis of the argument agrees with Stroud's (1968) seminal discussion. Stroud thought that Kant's argument had a much

better chance of defending the necessity of using certain concepts than later transcendental arguments because of its generality. Its claim is not just that some concepts must be used if others are—but that the possibility of thinking itself implies the necessity of using such concepts.

2. The arguments for mental act-consciousness and mental unity in the TD

The TD proper, the second chapter of the Analytic of Concepts, does not mention all the categories, much less argue for them. Rather, it tries to establish two preliminary results that are necessary for the eventual argument for the categories in the Principles chapter: the unity of the thinker and the necessary agreement between concepts and intuitions. I'm going to focus on the first task, that of establishing the unity of 'transcendental apperception'. In the *Critique*, Kant expressly distinguishes apperception from inner sense. We can get a better appreciation of his theory of apperception by briefly considering why he initially considered inner sense to be a key mental faculty. I begin with Locke's well-known introduction of inner sense.

> The other Fountain, from which Experience furnisheth the Understanding with *Ideas*, is the Perception of the Operations of our own Minds within us, as it is employ'd about the *Ideas* it has got; which Operations, when the soul comes to reflect on, and consider, do furnish the Understanding with another set of *Ideas*, which could not be had from things without: and such are *Perception, Thinking, Doubting, Believing, Reasoning, Knowing, Willing*, and all the different actings of our own Minds . . . (Locke 1690/1979: 105)

Although the notion of an internal sense was, I think, novel, one phenomenon that underlies its introduction was familiar. Many in the logical tradition maintained that people are aware of the cognitive acts they perform. Taking the most prominent example, the Port Royal Logic (Arnauld 1662) assumes that anyone can tell when he is judging, inferring, remembering, seeing, and so forth.

Since this phenomenon is much less remarked today, I will try to make it vivid with an example. Consider the premises of a simple inference:

> All men are mortal.
> Caius is a man.

Normal humans are aware both of the conclusion and of a movement of their minds from the premises to the conclusion. With no effort at all, they can distinguish acts of inferring from cases where they see no relation and have to be told what follows from the premises. This distinction would be obvious even if someone were to say the conclusion as quickly as the mind might infer it; it would be obvious even if a neurosurgeon could induce the production of sub-vocal speech, 'Caius is mortal', as quickly as the mind could infer that conclusion.

The phenomenon of mental act awareness justifies neither of Locke's assumptions. It implies neither that being so aware supplies thinkers with a concept of 'reasoning' nor that the awareness is best modelled by analogy with the 'outer' senses. I highlight mental act awareness not to support Locke's introduction of 'inner sense', but to draw attention to one phenomenon that it was meant to illuminate, a phenomenon that is central to Kant's theory of thinking.

For much of his career Kant was an inner sense enthusiast. He lauds the power of inner sense in an early essay (*The False Subtlety of the Four Syllogistic Figures*, 1762) where he explains the difference between so-called 'animal cognition' and rational human cognition. He is criticizing one of his contemporaries (the logician Meier) who had claimed that animals use concepts.

> [Meier's] argument runs like this: an ox's representation of its stall includes the clear representation of its characteristic mark of having a door; therefore, the ox has a distinct concept of its stall. It is easy to prevent the confusion here. The distinctness of a concept does not consist in the fact that that which is a characteristic mark of the thing is clearly represented, but rather in the fact that it is recognized [*erkannt*] as a characteristic of the thing. The door is something which does, it is true, belong to the stall and can serve as a characteristic mark of it. But only the being who forms the judgment: **this door belongs to this stable** has a distinct concept of the building, and that is certainly beyond the powers of animals.
>
> I would go still further and say: it is one thing to **differentiate** things from each other, and quite another thing **to recognize** [*erkennen*] the difference between them . . . (2.59–60)[4]

Animals can differentiate things from one another—in the sense that they can behave differently with respect to them. But that does not imply that they have any understanding of how they differentiate the objects.

[4] This translation is from (Walford and Meerbote 1992: 103–4).

The essay continues by offering a hypothesis about how humans are able to recognize characteristic marks as such and so have (distinct) concepts.

> My current opinion is that this power or capacity is nothing other than the faculty of inner sense, that is to say, the faculty of making [*zumachen*] one's own representations the objects of one's thought. This faculty cannot be derived from any other faculty. It is, in the strict sense of the term, a fundamental faculty, which in my opinion, can only belong to rational beings. But it is upon this faculty that the entire higher faculty of cognition is based . . . (2.60, Walford and Meerbote 1992: 104)

Why couldn't an animal just recognize the mark as such, why must it be able to think about its own representations? Kant's view of rational cognition is that in applying concepts, rational animals know the basis or ground for the application—hence they must be aware of their own representations, because those are the grounds of the application. By contrast, animals differentiate things only 'blindly', without any idea of the basis of their differential behaviour. I stress this point, because the cognition that is the central topic of the First Critique is the rational cognition just described.[5] Although it sounds somewhat oxymoronic, Kant's quarry is thus rational empirical cognition.

Johann Nicolaus Tetens wondered how the representations of inner sense could be understood as *representations* in the same sense as representations of outer sense (Tetens 1777/1979: 1.7.45). He answers his own question as follows: as objects cause impressions on sensory organs that give rise to sensations that represent the objects, (mental) acts that result in changes in representations cause impressions on the organ, the mind or brain, and those impressions give rise to sensations—which represent the actions. In light of his analysis, Tetens accepts a criticism of the *cogito* that he attributes to Johan Bernard Merian (1732–1807): Descartes should not have said 'I think', but 'I have thought' (Tetens 1777/1979: 1.47). Below I suggest why Kant might have demoted inner sense—as Tetens explained it—to a 'lower' faculty.

Let us turn to the TD itself or at least to some pieces of it. In the A Deduction, Kant offers a complex account of the various acts of putting representations together—or synthesizing them—that are necessary for empirical cognition. The necessity of the unity of apperception is introduced in the course of his account of the third synthesis, that of recognition in a

[5] I argue for this important point in *Kant's Thinker* (2011: Ch. 9).

concept. Unlike most discussions, he uses an example to explore the requirements of concept application. The example is counting:

> Without the consciousness that what we are thinking is the same as what we thought an instant before, all reproduction in the series of representations would be futile . . . If, in counting, I forget that the units that now float before my mind or senses were added together by me one after another, I should never know that a total is being produced through this successive addition of unit to unit, and so would remain ignorant of the number . . . (A103, amended translation)

As has often been noted, he believes that concept application involves the use of rules. In this case, the person must be aware of applying the counting rule to the units (perhaps a real or imagined stroke symbol): the first is designated by '1', etc. But something further is required. To have rational cognition—to know the basis of his judgement 'four', for example—the counter must recognize that judgement as the result of his application of the counting rule to his sensory evidence. How does this happen?

In the further discussion of the example, Kant suggests that a cognizer must be conscious of his act of judging:

> This number's concept consists solely in the consciousness of this unity of synthesis.
>
> The very word 'concept' could on its own lead us to this observation. For this one consciousness is what unites in one representation what is manifold, intuited little by little, and then also reproduced. Often this consciousness may be only faint, so that we do not [notice it] in the act itself, i.e. do not connect it directly with the representation's production, but [notice it] only in the act's effect. Yet, despite these differences, a consciousness must always be encountered, even if it lacks striking clarity; *without this consciousness, concepts, and along with them cognition of objects, are quite impossible.* (A103–4, my emphasis)

This is a strong claim. Cognition of objects would be impossible without a consciousness of acts of combining. But it should be clear why he thinks that these acts must be conscious. If cognizers were not conscious of these acts, then they would not know the basis of their judgements, and so would fail to be (rational) cognizers. He will allow that thinkers do not have to pay much attention to individual steps, adding up the stroke symbols little by little in accord with the counting rule; still they must be conscious that the act of judging 'four' is based on carrying out these steps.

For Kant, rational cognition requires a kind of act-consciousness. In his published Anthropology lectures, he characterizes consciousness of mental acts in terms of 'apperception' and relates that faculty to the understanding;

he contrasts 'apperception' with 'apprehension', a consciousness of particular mental states through inner sense (7.134n.). We can get a firmer grasp on Kantian apperception if we consider why he would have rejected inner sense, as Tetens understood it, as the basis of rational cognition. According to Tetens, inner sense is a sense, because it records acts of thinking. But mere awareness that you have judged could not underpin the rationality of the judgement. For that, the cognizer must be conscious, not that he has judged or even that he is judging right now. He must be conscious of judging on the basis of evidence, of having applied the rule to the data. Failing this, rational cognition through concepts—judging—would be impossible.

Consider also the 'Caius is mortal' inference. On Tetens' theory, a cognizer would know that he had inferred by being aware of the trace left by his act of inferring 'Caius is mortal'. Even supposing that inferring and other mental acts have somewhat different 'feels' or 'flavours', so that a reasoner can tell that 'Caius is mortal' was an inference and not a perception or something learned through testimony, this would hardly be sufficient to make him a rational reasoner. To be capable of rational inference, the reasoner must be aware—as he makes the inference—of his act as being based on premises. The creation of impressions of mental actions in a Tetensian inner sense is too little and too late to contribute to the rationality of inference or judgement.

In the counting passage, Kant has the cognizer calling the acts of applying the rule to units 'mine'. What is the justification for saying that the numbers were added together by *me*, for example?[6] Kant later agrees with Hume that humans can see no constant self in the flux of mental states:

> There is, in inner perception, a consciousness of oneself in terms of the determinations of one's state. This consciousness of oneself is merely empirical and always mutable; it can give us no constant or enduring self in this flow of inner appearances. (A107)

He also disagrees with Locke that the mere consciousness of states can explain how different states belong to a single 'I'. In the B edition of the

[6] Kant's explanation concerns why different states should be understood as belonging to a common subject—what we might call the 'togetherness' of different states. He does not address (except inadequately through inner sense) why these states should be understood as belonging to a particular subject, namely me. That is, he does not adequately address what might be called the 'mineness' problem of mental states. See (Kitcher 2011: Chs. 1, 2, 9, and 15) for further discussion.

TD he is explicit about the problem with Locke's theory (though he doesn't mention Locke by name):

> The empirical consciousness that accompanies different representation is intrinsically sporadic and without any reference to the subject's identity. (B133)

Kant's objection is that Lockean consciousness—the consciousness which is inseparable from thinking (Locke 1690/1979: 335)—is momentary or episodic. As such, it cannot provide a basis for representing a common subject.

So how can humans be aware of themselves as continuing, existing, thinkers? The key is their awareness of their acts of synthesizing:

> Reference to the subject's identity . . . comes about not through my merely accompanying each representation with consciousness, but through my adding one representation to another *and being conscious of their synthesis*. Hence only because I can combine a manifold of given representations **in one consciousness**, is it possible for me to represent the **identity of the consciousness itself in these representations**. (B133, my emphasis)

But how does being aware of these acts of synthesis enable the cognizer to represent his identity?

Kant takes the representation 'I think' to be *a priori* (B132). We have just seen the argument that it cannot be *a posteriori*: neither inner intuition nor a Lockean 'accompanying' consciousness (if those are different) provides any evidence of a continuing self. As with the *a priori* categorial concepts, the 'I think' is associated with an *a priori* principle, the principle that different representations must belong to a common self or thinker (A117, B132). No amount of empirical data could establish this principle. It is nonetheless possible for cognizers to recognize instances of it. To see how and also how consciousness of synthesis permits subjects to recognize their identity, we may return to the counting example.

The counter is aware of four stroke symbols to which he applies the counting rule, 1, 2, etc. When the understanding applies the counting rule to information contained in the sensory states that float before the mind, it recognizes that the antecedent of the rule is fulfilled, so the judgement '4' can be made. But it also recognizes something else. Through being aware of its act of synthesis, it recognizes that it has made the judgement on the basis of applying the counting rule to representations contained in sensory states. Thus, it recognizes that the judgemental state could not exist without the sensory states. The judgemental state must belong with the

sensory states to a single consciousness. That is, the understanding recognizes that the mental states it combines and the combined state that results from the combination as instances of the 'I think' rule. Because a counter applies two rules, the counting rule and the rule of apperception, she does not merely form the representation '4', she also represents the states and acts of which she is conscious as the states and acts of a single cognizer. Consciousness of the act of synthesis is crucial for rational cognition; without it, conceptual or rational cognition of objects is impossible. With it she is also able to recognize the unity of her consciousness. This is why Kant claims in the 'anti-Locke' passage that it is *only* through engaging in cognition that one can recognize the identity of one's consciousness through time. He repeats this extraordinary claim in the B edition discussion of the Paralogisms:

> We are acquainted with the unity of consciousness itself only by its being for us an indispensable requirement for the possibility of experience. (B420)

The B Deduction makes an even stronger claim about the relation between cognition and the unity of consciousness.

> Synthetic unity of the manifold of intuitions, as given *a priori*, is thus the ground of the identity of apperception itself. (B134)

It is not just that, if the multifarious contents of sensory states could not be combined in judgements, then would-be cognizers would lack experience (empirical cognition). Nor is the claim even that, under these circumstances, they would not be able to *think* about themselves as continuing cognizers. It is rather that, absent the combinability of intuitions, the identity of apperception would also be absent. As he explains slightly later:

> All representations given to me must stand under this [original synthetic unity of apperception], however they must be brought under it through a synthesis. (B136)

That is, the unity of self-consciousness is brought about through combination. In engaging in cognition, the understanding also partially creates a rational subject.[7]

[7] See also (A112): 'Without such unity [as produced by the rule for cause and effect, for example] no thoroughgoing and universal and hence necessary unity of consciousness would be encountered in the manifold of perceptions.'

I hope that I have laid out enough of the argument of TD that its structure and principal theses are clear. The argument regresses from the possibility of rational empirical cognition to the unity of apperception; having shown how thinking must work for creatures like us who acquire knowledge through combining information gained through their senses, Kant is then in a position to consider what the possibility of thinking itself requires. And what it requires is that the materials received through the senses can be combined in ways that make thinking and so the unity of apperception possible. He claims, but does not argue, that those ways of combining materials are captured in the *a priori* categories. Even at this point, however, we can see why he makes the strong and unusual claim that object cognition and the unity of apperception are necessary and sufficient conditions for each other (for example, B137). Failing the combination of some representations in a further representation such as a judgement or the conclusion of an inference, a cognizer would be unable to recognize the relation of necessary connection across her states that stands behind the use of 'I think'; failing the ability to apply the 'I think' across her representations, a cognizer would lack rational empirical cognition, because she would not see some of her representations as the grounds of her judgements.

3. Arguing for mental unity

We can see how strong an argument Kant is able to mount against those who deny the legitimacy of an *a priori* concept of 'person' or 'unified cognitive subject' by contrasting it with one of the best recent attempts at presenting a 'transcendental argument' for the same conclusion, that of Quassim Cassam (1997). The essentials of Cassam's wide-ranging account can be captured as follows. Neo-Kantians have argued that it is impossible to think of one's experience as containing objects in the weighty sense (items that can be perceived and that can exist unperceived) unless one can self-ascribe perceptions and grasp the identity of the thing to which these perceptions are ascribed (Cassam 1997: 36). Because of the resemblance of this line of reasoning to Kant's argument from object cognition to the unity of consciousness, Cassam refers to it as the 'objectivity requires unity' (ORU) argument. He notes that those who argue that there are no continuing persons will not be impressed. Theirs is a thesis about what

those objects who were thought to be persons really are; ORU concerns the ways in which individuals must think of themselves in having cognition (Cassam 1997: 178–9, 181). He presses on, because he thinks that, thanks to ORU, there is something to be explained—namely, how cognizers use the term 'I'—that may require reference to persons.

At this point, Cassam appeals to Gareth Evans' (1982) analysis of the range of capacities that are required for individuals to have 'I thoughts': such individuals must be able to recognize the connection between 'I thoughts' and their special ways of gaining knowledge of their mental states and physical properties, must recognize the connection between 'I thoughts' and behaviour, and so on (Cassam 1997: 189). So the argument is that objectivity requires unity, including the possession of 'I thoughts' and—moving from epistemological considerations to the grounding of cognitive capacities in a subject—'I thoughts' can be had only by creatures with various further capacities, who are thus persons, or substantial subjects among other items in the (physical) world (Cassam 1997: 196–7). As Cassam concedes, this argument is weaker than it might be. It avoids the fallacy of arguing from what cognizers must think they are to what they are, by making a large assumption: only substantial subjects can have the capacities required to be thinkers of 'I thoughts' (Cassam 1997: 197).

Kant's version of the argument from (rational empirical) cognition to unity of consciousness includes an element that is missing from the neo-Kantian repertoire. To return, again, to the counting example, a counter is aware of four stroke symbols to which he applies the counting rule, 1, 2, etc. When he applies the rule, he recognizes that the antecedent of the rule is fulfilled, and so judges '4'. Because he is at least implicitly aware of the act of synthesis, he recognizes that he has made the judgement on the basis of applying the counting rule to representations contained in sensory states—and that his act of judging thereby creates a relation of necessary connection across the sensory states and the judgemental state. Through the self-conscious act of judging, the judgemental state comes to stand in the relation of rational dependence to the sensory representations; from the other direction, through that act, mental states of ticking off '1', '2', etc., achieve the status of grounds of cognition through the rational dependence of the judgemental state upon them. Consciousness of the act of synthesis is crucial for rational cognition. Without it, conceptual or rational cognition of objects is impossible, because cognizers would not know the bases of their cognitions. With that consciousness, however, the

cognizer creates a relation of rational dependence across his states in part by being at least implicitly cognizant of that relation. Since cognizers come with an *a priori* representation 'I think' that they apply according to the rule of necessarily belonging together, they can always attach 'I think' to the representations that participate in rational cognition.

Perhaps the view I am attributing to Kant will be clearer if we return to Stroud's classic criticism of contemporary transcendental arguments. In Stroud's view (1994: 234),[8] Kant tried to legitimate the categories by arguing that they were both indispensably necessary for empirical cognition and objectively valid—they were true of objects of experience. The fundamental error of the neo-Kantians was to assume that the first project sufficed for the second. It could not, because the first project concerned only what cognizers must believe (and so what concepts they must possess) and there is no legitimate way to argue from the necessity of a belief in Xs to the objectivity validity of the concept 'X'. Stroud's criticism is widely recognized as sound and has shaped many subsequent discussions, including Cassam's.[9] He was also right about Kant's intentions. In the case of the representation 'I think', the aim was not just to show that this *a priori* representation was necessary for rational empirical cognition, but that rational cognition required the existence of thinkers whose states were related as specified by that representation. And that is what Kant's theory implies. Rational cognition does not just require that cognizers have an *a priori* representation 'I think'. In any case of rational cognition, a cognizer must grasp, in making the judgement, that the relations of epistemic dependence specified in the rule attached to the 'I think' are fulfilled. A judgement is rational only when it has that genesis. Hence the *a priori* representation 'I think' and its associated rule play an essential role in producing rational judgements and thus rational judgers or thinkers.

In this respect, the representation 'I think' is unique. Having a concept of external objects or of other minds is neither a necessary nor a sufficient condition for the existence of external objects or of other minds. But—when given suitable materials to work on and mental act awareness—the concept of a thinker whose representations stand in relations of necessary connection is a necessary and sufficient condition for a cognizer to form a rational judgement and so to be a thinker. In this case alone, the

[8] See also (Stroud 1968).

[9] Cassam (1987) discusses Stroud's argument at some length. See also his (2003).

representation (plus other representations and the mental activity of combining them) is able to create the reality.

To move beyond the world of belief and representation to that of reality, Cassam appeals to the fact that various mental capacities require the use of a body. The representation 'I think' required by cognition of objects can therefore be deployed only by embodied or substantial subjects. Oddly, Kant is able to move from representation to reality by staying within the realm of cognizing and thinking.

4. Consciousness of mental acts[10]

One important part of Kant's account of rational cognition is the claim that cognizers must be conscious of their mental acts of judging and inferring. What kind of consciousness is this? As noted, Chalmers has distinguished easy kinds of consciousness, those that can be modelled, for example, by a computer, from the hard kinds, those that seem to resist modelling by any mechanism that is currently available. For many years the view has been that sentience, the ability to feel and sense, raises the hard problem.[11] By contrast, sapience, the ability to think and represent, was easy. After all, the first computer models of mental processes were devoted to proving logical theorems and solving word problems. Apparently, these tasks could be carried out by simple production systems, that is, sets of conditional rules that tell the machine what formula to write down depending on the program and the current state. On Kant's view, however, a simple production system could not model rational cognition.

Kant maintains that such cognition requires mental act awareness. It is tempting to think that a familiar complication to simple production systems could handle the needed awareness. Besides carrying out various rewrite rules, the program would contain a function that monitored its

[10] Recently Christopher Peacocke (2008: Ch. 7) has started to investigate mental act-consciousness. His approach is to model mental act-consciousness on physical act-consciousness. *Prima facie* the cases seem disanalogous, because mental act-consciousness is essential to the performance of certain kinds of mental acts, such as judging and inferring. This doesn't seem to be the case with physical actions, unless a physical motion can be understood as 'action' only when embedded in a rich description of mental activity. In that case, however, what needs to be understood is the mental activity.

[11] Some, notably John Searle (1980), have argued that sapience is also a hard problem, but the 'standard' view is that the hard problem arises mainly in the case of sentience.

own rewriting activities. To evaluate this suggestion, I turn to Ned Block's (1995) account of models of monitoring consciousness. Block offered three ways of modelling monitoring consciousness: (1) as metacognition, where being in a mental state is accompanied by the thought that one is in that state; (2) as self-scanning; and (3) as a phenomenal consciousness, where one is aware of some sort of phenomenal quality. Perhaps in this case it would be the feeling of judging.

None of these models seems to fit Kant's description of judging or concept application. We have already seen the difficulty with (1). Rational cognition is not simply a matter of knowing that you are in a state, say, of judging, but of being conscious of entering that state through a synthetic act. The second option seems equally unpromising. If a sequence of mental states did not involve an awareness of entering the latter state through a synthetic act, then scanning that sequence cannot make up for the lack of this awareness. The third option seems even less plausible. Even if we assume that different mental actions have different feels, judging and inferring do not seem to be centred on phenomenal consciousness in the way that the paradigm cases of feeling pain or seeing purple are. If there is a phenomenal feeling to these acts, it is not their essence, which concerns the ability to recognize the relations across one's own thoughts.

One symptom that a type of consciousness falls in the 'hard' category is the susceptibility of putative accounts to 'Zombie' objections. Thinking about Zombies (those who act just like people, but lack consciousness) may make it clearer why Kant downgraded 'inner sense' (on Tetens' theory of what it involves). Insofar as Tetens is right about how inner sense works, then inner sense is not sufficient for explaining rational cognition. Creatures with just inner sense would be cognitive Zombies. They would look and sound quite like cognizers, but they would not be rational cognizers. Such creatures would process information or, in Kant's terminology, combine representations in regular ways. They might also have metacognitive beliefs about the information (or representations) that they have received. They might be aware that some of their representations have the feel of judgements, so they could report which of their beliefs come from their own processing or combining as opposed to being received from others. Finally, they might even be aware of which of their representations are connected to their judgements as their bases. In Kant's view, even if such Tetensian creatures could say which of their representations were judgements and could connect these to their bases, they would

still not be rational cognizers, because they could not make the connection in the right way. They could not see their judgements as the results of acts of judging on the basis of the perceptual evidence. Alternatively, although they would know in a sense that these are judgements that they made rather than received, they would not see any fundamental difference between thinking for themselves and taking on the opinions of others. Still a third formulation: although they would connect judgements to their bases, the connections would seem accidental, because they would not understand how the judgemental representations were dependent on the earlier representations. The problem with modelling Kant's mental act awareness as monitoring consciousness is that the latter is a contemporary reincarnation of Locke's reflective consciousness, the consciousness that does or can accompany mental states. Kant saw clearly that an episosdic reflecting consciousness could not account for mental unity, the relations of necessary connection across mental states. Since reflecting or monitoring consciousness could not account for the relations of epistemic dependence across the states of a rational cognizer, it could not explain rational cognition either. These intertwined inadequacies of monitoring consciousness are just what led Kant to introduce a new kind of awareness, mental act-awareness or apperceptive consciousness.

References

Arnauld, A. 1662. *The Art of Thinking*. Translated by J. Dickoff and P. James. Indianapolis: Bobbs-Merrill.

Block, N. 1995. A Confusion about a Function of Consciousness. *Behavioral and Brain Sciences* 18: 227–87.

Carl, W. 1989. Kant's First Drafts of the Deduction of the Categories. In E. Förster (ed.), *Kant's Deductions*. Stanford: Stanford University Press, pp. 3–20.

—— 1992. *Die Transzendentale Deduktion der Kategorien in der ersten Auflage der Kritik der reinen Vernunft: Ein Kommentar*. Frankfurt am Main: Vittorio Klostermann.

Cassam, Q. 1987. Transcendental Arguments, Transcendental Synthesis, and Transcendental Idealism. *Philosophical Quarterly* 37: 355–78.

—— 1997. *Self and World*. Oxford: Oxford University Press.

—— 2003. Can Transcendental Arguments be Naturalized? *Philosophy* 78: 181–203.

Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

Evans, G. 1982. *The Varieties of Reference*. Edited by J. McDowell. Oxford: Oxford University Press.

Guyer, P. and Wood, A. W. (Trans.) 1998. *The Critique of Pure Reason. The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

Henrich, D. 1989. Kant's Notion of a Deduction and the Methodological Background of the First Critique. In E. Förster (ed.), *Kant's Deductions*. Stanford: Stanford University Press, pp. 229–46.

Kant, I. 1900–. *Kants gesammelte Schriften, Akademie Ausgabe*. Edited by the Koeniglichen Preussischen Akademie der Wissenschaften, 29 vols. Walter de Gruyter and predecessors.

Kemp Smith, N. 1968. *Immanuel Kant's Critique of Pure Reason*. London: Macmillan.

Kitcher, P. 2008. Kant's I think. In V. Rohden, R. R. Terra, and G. A. de Almeida (eds.), *Recht und Freiden in der Philosophie Kants: Akten des X Kant-Kongresses*. New York: Walter de Gruyter.

—— 2011. *Kant's Thinker*. Oxford: Oxford University Press.

Kuehn, M. 1997. The Wolffian Background to Kant's Transcendental Deduction. In P. A. Easton (ed.), *Logic and the Workings of the Mind*. Atascadero, CA: Ridgeview, pp. 229–50.

Locke, J. 1690/1979. *Essay Concerning Human Understanding*. Edited by P. H. Nidditch. Oxford: Clarendon Press, 1979.

Peacocke, C. 2008. *Truly Understood*. Oxford: Oxford University Press.

Pluhar, W. (Ed. and trans.) 1996. *Critique of Pure Reason: Unified Edition*. Indianapolis: Hackett.

Searle, J. 1980. Minds, Brains, and Programs. *Behavioral and Brain Sciences* 3: 417–24.

Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241–56.

—— 1994. Kantian Arguments, Conceptual Capacities, and Invulnerability. In P. Parrini (ed.), *Kant and Contemporary Epistemology*. Dordrecht: Kluwer, pp. 231–52.

Tetens, J. N. 1777/1979. *Philosophische Versuche über die menschliche Natur und ihre Entwicklung*, 2 vols. Leipzig: M. G. Weidmans Erben und Reich. Reprinted by the Kantgesellschaft Verlag.

Walford, D. and Meerbote, R. (Eds.) 1992. *Kant's Theoretical Philosophy, 1755–1770. The Cambridge Edition of the Works of Immanuel Kant*. Cambridge: Cambridge University Press.

Wilson, M. 1974. Kant and the Refutation of Subjectivism. In L. W. Beck (ed.), *Kant's Theory of Knowledge*. Dordrecht: Reidel, pp. 306–16.

Wolff, C. 1751/1983. *Vernuenftige Gedancken von Gott, der Welt under der Seele des Menschen, auch allen dingen ueberhaupt*. Introduction (in English) by C. A. Corr. Hildesheim: Georg Olms Verlag.

5

The Value of Humanity: Reflections on Korsgaard's Transcendental Argument

Robert Stern

The purpose of this chapter is not to consider the worth of transcendental arguments in general (which I have done at length elsewhere),[1] but instead to focus on a specific example of the genre. However, this is an example taken not from epistemology or metaphysics where (in Anglo-American philosophy at least)[2] such arguments have most usually found a home, but rather from ethics. Nonetheless, I hope that what I have chosen to do will not prove valueless, as although specific and in some ways untypical, the argument I set out to discuss is important and influential, but not much considered as a transcendental argument as such. The argument in question is Christine Korsgaard's, from her book *The Sources of Normativity*. My aim is to try to understand what role her transcendental argument for the value of humanity is meant to play in her project, and whether the argument succeeds. Rather to my surprise, and rather against the run of the critical literature on Korsgaard's book, I will suggest that in one of its forms, the argument can be made to work—at least in its own terms[3] and when its rather limited place in Korsgaard's overall strategy is understood.

[1] Most particularly in (Stern 2000).

[2] Transcendental arguments have of course been used by Habermas, Apel, and others as part of their projects in social philosophy. For a discussion of current uses of transcendental arguments in ethics, but which curiously hardly mentions Korsgaard, see (Illies 2003).

[3] For example, as we shall see further below, I think the argument relies heavily on other arguments Korsgaard gives against realism, where in this discussion I allow her to take these for granted (but where I am critical of them elsewhere).

And in the end, of course, it may also turn out that there are general lessons to be gained from the examination of this argument after all.

I

Taken as a whole, Korsgaard's aim is to show that we stand under moral obligations, by constructing an argument to that effect. But only part of the overall argument is meant to be a transcendental one—roughly, the middle part. How that specifically transcendental argument is understood therefore depends on how one conceives of what precedes it, and what work its conclusion is supposed to do in what follows. Let me begin by setting out in broad terms what I take the three phases of Korsgaard's overall argument to be.

Phase One: From free agency to the categorical imperative

Korsgaard starts from a Kantian antinomy: that we conceive of ourselves as free, on the one hand, but also as agents with a certain stability of purpose and character, on the other, which means we take our actions to be governed by principles or laws. Korsgaard's Kantian solution to this antinomy is to argue that these principles or laws must be self-imposed.[4] But now we seem to face a second antinomy: on the one hand, unless a principle or law already binds the will, on what basis can the will rationally legislate a law to itself; but on the other hand, if the will is already bound by a law, how can this count as self-legislation?[5] Korsgaard's proposal is that Kant's notion of the categorical imperative is designed to resolve this second antinomy: on the one hand, the free will is not completely lawless,

[4] Cf. (Korsgaard 1996b: 97–8): '[Kant] defines a free will as a rational causality which is effective without being determined by any alien cause . . . including the desires and inclinations of the person. The free will must be entirely self-determining. Yet, because the will is a causality, it must act according to some law or other. Kant says: "Since the concept of a causality entails that of laws . . . it follows that freedom is by no means lawless . . . " Alternatively, we may say that since the will is practical reason, it cannot be conceived as acting and choosing for no reason. Since reasons are derived from principles, the free will must have a principle. But because the will is free, no law or principle can be imposed on it from the outside. Kant concludes that the will must be autonomous: that is, it must have its *own* law or principle.'

[5] Cf. (Korsgaard 1996b: 98): 'And here again we arrive at the problem. For where is this law to come from? If it is imposed on the will from outside then the will is not free. So the will must make the law for itself. But until the will has a law or principle, there is nothing from which it can derive a reason. So how can it have any reason for making one law rather than another?'

because it must act in accordance with a law or principle; but on the other hand, this does not constrain it or make it less free, because it is just constitutive of free legislation that it has this structure,[6] while which law it chooses is left open.[7] The conclusion of this first phase of the argument is therefore that in order for the will to be free, it must act on the basis of something that has the nature of a law or principle.

Before moving on to Phase Two of Korsgaard's overall argument, it may be worth pausing to emphasize that (at least as I see it) Phase One is not a transcendental argument: rather, it is an argument that works by showing how the antinomies of free agency and of self-legislation need to be resolved, leading to the categorical imperative, to act only on a maxim that I can will as a universal law.

Phase Two: From the categorical imperative to the moral law, step 1: the value of your own humanity

In a way that she represents as a departure from Kant,[8] Korsgaard says that the strategy of Phase One cannot in itself take us as far as the *moral* law: that is, it cannot establish that the law we must abide by is one that constrains our treatment of others in any recognizably moral way, in either a positive or negative sense, in terms of our obligations to do things for them, or to avoid acting against them. For example, the rational egoist acts on a practical law, inasmuch as she adopts the principle of always acting to promote her interests, and that seems sufficient to provide the kind of coherent structure to the will that is constitutive of free agency on this

[6] Cf. (Korsgaard 1996b: 235): 'Cohen makes it sound as if autonomous lawmaking were one thing, and universal autonomous lawmaking another, and this in turn makes it sound as if universalizability is a rational constraint which is imposed on what would otherwise be the arbitrary or unconstrained activity of autonomous lawmaking. But I think Kant himself means something else, namely autonomous lawmaking just *isn't* autonomous lawmaking unless it is done universally. The requirement of universalization is not imposed on the activity of autonomous lawmaking by reason from outside, but is constitutive of the activity itself.'

[7] Cf. (Korsgaard 1996b: 98): 'The problem faced by the free will is this: the will must have a law, but because the will is free, it must be its own law. And nothing determines what that law must be. *All that it has to be is a law.* Now consider the content of the categorical imperative, as represented by the Formula of Universal Law. The categorical imperative merely tells us to choose a law. Its only constraint on our choice is that it has the form of a law. And nothing determines what the law must be. *All that it has to be is a law.*'

[8] Cf. (Korsgaard 1996b*:* 98–100, 221–2, 233, 237). As she also points out (e.g., p. 99, n.9), this is also a departure from her earlier self in papers reprinted in (Korsgaard 1996a), where she moves straight from her solution to the Kantian antinomy of self-legislation to the moral law: cf. (pp. 166–7).

account. Korsgaard therefore allows (in a way that a more traditional Kantian might not)[9] that nothing she has said so far establishes that the principle the agent needs to adopt is anything we would recognize as a moral principle, 'the law of what Kant calls the Kingdom of Ends, the republic of all rational beings' (Korsgaard 1996b: 99). To get to *this* law, Korsgaard holds, she must first argue that you must place a value on your own humanity (which is what she does in Phase Two), and then that you must value the humanity of others (which is what she does in Phase Three); once this is established, then the agent cannot adopt the principle of self-interest or any other such non-moral principle as her law, because this would violate the value of humanity. As Korsgaard puts it: '[The] argument . . . aims to move from the formal version of the categorical imperative to moral requirements by way of the Formula of Humanity' (Korsgaard 1999: 28, n.23).

On this approach, then, the role of Phase Two is to be a stepping stone to establishing the value of humanity, which is a conclusion of the whole argument in Phase Three; and that stepping stone is to establish to the agent the value of her humanity. To see why Phase Three is required, we can ask how Phase Two falls short: why isn't establishing the value of the agent's own humanity sufficient to lead the agent to adopt the moral law? The answer, of course, is that even if the agent recognizes the value of her own humanity, to be moral she needs to respect the value of others, and it is this shift from agent-relative to agent-neutral reasons that Phase Three is designed to achieve.

Phase Three: From the categorical imperative to the moral law, step 2: the value of humanity in general

As with Phase One, this phase of the argument is not a transcendental one;[10] and it is perhaps the part of Korsgaard's overall approach that has been most brusquely dismissed by critics. Korsgaard's strategy here is to claim that the egoist's agent-relative reasons are private, and then to use considerations from Wittgenstein's private language argument to show that this would make them incoherent as reasons, for this argument shows

[9] Cf. (Ginsborg 1998: 8).

[10] At least, I don't think it is, and Korsgaard doesn't claim it is either. But for the suggestion it should be seen in this way, see (Skidmore 2002: 135).

that reasons must be public to be reasons at all.[11] But Korsgaard's critics have been unimpressed by the suggestion that agent-relative reasons are in fact private ones, in any sense that brings in Wittgensteinian considerations: as Skorupski puts it, 'others can "share" the normative force of the egoist's reasons; that is, they can understand his reasons and, if egoism were right and they were rational, could acknowledge their force (as agent-relative reasons)' (Skorupski 1998: 348–9).[12]

This completes my outline of Korsgaard's overall argument. I will not say anything more about Phase One or Phase Three, and in particular I will not attempt to defend Korsgaard from her critics over the latter, because my focus here is intended to be on the transcendental part of her strategy, which is in Phase Two. It is therefore now time to look in more detail at what the transcendental argument in Phase Two is meant to be.

II

On the account I have given of Phase Two, the aim here is to establish the value of your humanity, as a way of moving to the establishment of the value of humanity in general in Phase Three, in order to show why the law that the agent chooses in Phase One cannot violate the dignity of persons.

To those familiar with debates surrounding transcendental arguments, this may immediately raise concerns. For, it may seem that Korsgaard is straightaway making claims for her transcendental argument that have been famously rendered problematic by Barry Stroud, by using that argument to establish a conclusion about how things are, viz. that your humanity has value. As has been much discussed, Stroud suggested in his 1968 article that such world-directed claims can invariably be resisted by the sceptic, and weakened to appearance or belief-directed ones, so that the conclusion of a plausible transcendental argument will only tell us how things must appear to us or how we must believe them to be, in order to make possible thought, language, experience or whatever.[13] If Korsgaard's transcendental argument is making a world-directed claim, therefore, it may seem that this Stroudian worry needs to be addressed.

[11] Cf. (Korsgaard 1996b: 132ff.).

[12] Skorupski is here echoing Nagel's complaint in his response to Korsgaard in (Nagel 1996: 208); and cf. also (Skidmore 2002: 135–7).

[13] (Stroud 1968: 255–6). For further discussion of Stroud's position, see (Stern 2000).

Of course, if Korsgaard wanted to defend a world-directed transcendental argument, she could perhaps do so by questioning Stroud's reasons for thinking that such arguments can always be weakened by the sceptic to what we must believe or how things must appear. But even if Stroud's position is questionable in general,[14] it may seem that there is something especially problematic in taking Korsgaard's argument in a strong or ambitious form: for it may just seem incredible to think that you could be given an argument to establish that your humanity has value, in a world-directed sense. This incredibility is clearly felt by Skorupski when he writes: 'It would be gratifying to have it demonstrated by pure philosophy that one is important . . . But in the absence of contagious magic the demonstration seems less than cogent' (Skorupski 1998: 350). The worry here, I think, is hubris: how can it be established that we have value as such, when seen in the scheme of things it seems we have no more significance than anything else—when, as Hume put it, 'the life of man is of no greater importance to the universe than that of an oyster' (Hume 1965: 301). Thus, even if a case could be made for world-directed transcendental arguments in general (contra Stroud et al.), is it reasonable to think that such a case can be made concerning my value as a human being?

However, whatever the force of these concerns, it is not clear that they are worries that need apply to Korsgaard. For she herself does not propose any such ambitious, world-directed transcendental argument,[15] but instead puts forward a modest argument, combined with a kind of anti-realism or perspectivism about value, that does get to the stronger conclusion that your humanity has value, but only where that value is conceived of in this perspectival way.[16]

[14] In (Stern 2007b), I argue that Stroud's argument is not as compelling as is usually assumed, but that a better argument can be offered to the same modest effect.

[15] I would therefore contrast Korsgaard's position with Allen Wood's, who follows a similar argument to Korsgaard, but to a more realist conclusion, but without attempting to address Stroudian concerns: cf. (Wood 1999: 125–32); and his review of Korsgaard's *Creating the Kingdom of Ends* in (Wood 1998).

[16] Skidmore accepts this defence of Korsgaard's transcendental argument in Phase Two, but as we have seen, thinks Korsgaard also has a transcendental argument in Phase Three, which is more ambitious, and so which falls foul of Stroud's criticisms: see (Skidmore 2002: 134–40). But as I have mentioned, Korsgaard herself doesn't present Phase Three as a transcendental argument, so it is debatable whether these issues apply to it. In her more recent rehearsal of the argument in (Korsgaard 2009: 22–5), she also distinguishes clearly between these phases, and again only talks of the second in transcendental terms.

That Korsgaard sees herself as approaching things this way is clear from what she says when she summarizes her transcendental argument:

> The argument I have just given is a transcendental argument. I might bring that out more clearly by putting it this way: rational action exists, so we know it is possible. How is it possible? And then by the course of reflections in which we have just engaged, I show you that rational action is possible only if human beings find their own humanity to be valuable. But rational action is possible, and we are the human beings in question. Therefore we find ourselves to be valuable. Therefore, of course, we are valuable. (Korsgaard 1996b: 123–4)

Korsgaard then goes on:

> You might want to protest against that last step. How do we get from the fact that we find ourselves to be valuable to the conclusion that we are valuable? (Korsgaard 1996b: 124)

And here is Korsgaard's response to this worry:

> [T]here's a good reason why the argument must take this form after all. Value, like freedom, is only directly accessible from within the standpoint of reflective consciousness. And I am now talking about it externally, for I am describing the nature of the consciousness that gives rise to the perception of value. From this external, third-person perspective, all we can say is that when we are in the first-person perspective we find ourselves to be valuable, rather than simply that we are valuable. There is nothing surprising in this. Trying to actually see the value of humanity from the third-person perspective is like trying to see the colours someone sees by cracking open his skull. From the outside, all we can say is why he sees them.
>
> Suppose you are now tempted once more to say that this shows that value is unreal just as colour is unreal. We do not need to posit the existence of colours to give scientific explanations of why we see them. Then the answer will be the same as before. The Scientific World View is no substitute for human life. If you think it is unreal, go and look at a painting by Bellini or Olitski, and you will change your mind. If you think reasons and values are unreal, go and make a choice, and you will change your mind. (Korsgaard 1996b: 124–5)

Korsgaard is thus agreeing with the central Humean idea, that from the point of view of the universe nothing really has value; but Korsgaard doesn't claim to be operating from that point of view. Rather, she is engaging with agents who have a perspective on the universe that involves the experience of values and making judgements about them, just as much as they have a perspective that involves experiencing colours and making judgements about them too. So, she holds, if we can establish that from

our perspective we must experience or judge that we ourselves have value, that is good enough for this exercise. Korsgaard can therefore agree with the Humean point that to think that we could establish anything more about this world is absurd, and can likewise claim that her transcendental argument doesn't have to be ambitious in this sense, while still insisting that no more than this is required, once this conception of value is accepted.

Now, of course, this approach to the notion of value, and whether it is substantive or realist enough, will be controversial;[17] but I take it here that enough people will find it congenial to serve as an adequate way of allowing Korsgaard to present her transcendental argument in modest terms.[18] We are now in a position to see what that modest transcendental argument is meant to be, where all that it intends to establish qua transcendental argument is that we must value our own humanity. I will offer two accounts of this argument, and claim that the second is to be preferred.

[17] In his response to Korsgaard in (Korsgaard 1996b), G. A. Cohen expresses some misgivings on this score: see Cohen (1996: 186); and for a defence of Korsgaard's approach, see (Gibbard 1999: 153): 'One aspect of Korsgaard's argument will be controversial, but not with me. It is transcendental: it takes something we can't act without accepting, derives a consequence, and then embraces the consequence . . . [But] if the argument goes through as intended, its conclusion doesn't follow logically from its premises—that's the worry. Mightn't it be that although merely acting at all commits us to *thinking* that humanity has value, in fact it doesn't *have* value? Korsgaard, though, say I, has every right to rely on such arguments. Suppose she is right, and in settling whether to act, I've settled whether to believe humanity valuable. I'll then act and voice the conviction to which acting commits me: Humanity is valuable. What other conceivable access can I have, after all, to the question of whether humanity is valuable, but to reflect on what to do? The value of humanity or its lack isn't a feature of nonnatural space, glimpsed by intuition. Thinking humanity valuable, if Korsgaard is right, is an inseparable part of thinking what to do and why. Whether to think humanity valuable is just the question, whether to value humanity.' Cf. also (Bittner 1989: 24): 'Now actually we may disregard the difference between showing that moral demands are valid and showing that they must be considered valid. For the realization that given certain basic features of our lives we cannot help but acknowledge moral demands is tantamount to having their validity demonstrated to us.'

[18] Skorupski seems to think that even such a modest transcendental argument is as problematic as a more ambitious one, when the full quotation from him cited earlier runs: 'It would be gratifying to have it demonstrated by pure philosophy that one is important. *Or even—to put it with due Kantian caution—that one must take oneself to be*. But in the absence of contagious magic the demonstration seems less than cogent' (my emphasis). This is because Skorupski does not see why there might not be valuable things to be done independently of our having value (cf. (Skorupski 1998: 350): 'Even if humanity is worthless might there not sill be valuable things to be done?'), or how our having value can 'magically' confer value on other things. I think Korsgaard's response, however, would be to say that the alternative is equally 'magical'—namely, how can things have an intrinsic value in themselves? For this sort of worry concerning intrinsic value, cf. also (Street 2008).

III

The first account of the argument I will consider runs as follows:

1. You cannot act unless you can take some impulse to be a reason to act.
2. You cannot take some impulse to be a reason to act unless it conforms to some way in which you identify yourself (a practical identity).
3. You cannot adopt a particular practical identity unless you also adopt humanity as a practical identity.
4. You cannot adopt humanity as a practical identity unless you value your humanity.
5. Therefore, you must value your humanity, if you are to be an agent.

Let me consider in more detail what this all means.

As with most transcendental arguments, Korsgaard is trying to begin with a premise that the sceptic can be expected to accept, where here the premise is that he is capable of agency. Now, by this Korsgaard doesn't just mean behaviour or bodily movement, but some exercise of the will. Moreover, she holds, to exercise the will and so act in this way, it is not sufficient that whenever an impulse to act or a desire assails you, you will follow it, for then you are not deciding to act at all. Rather, the way to act is to act for a reason: to decide that this impulse (for example, to buy this toy) is a good one (for example, because it will make my daughter happy).

The second step is to introduce a more distinctively Korsgaardian idea: that if action requires the having of reasons, those reasons are not 'out there' in the world, but come from the way in which doing certain actions would relate to the kind of person you are. Thus, it is qua my daughter's father that I have a reason to buy her this toy, if it would make her happy. As Korsgaard puts it: 'It is necessary to have some conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining whether they are consistent with the ways in which we identify ourselves' (Korsgaard 1996b: 120). Unless we had some such practical identity, Korsgaard claims, there would be no reason for us to act on one impulse rather than not, and thus no possibility of rational agency at all.

Obviously one way to resist Korsgaard's argument at this point would be to opt for a more realist conception of reasons, and to claim that for us

to have a reason to act on an impulse is just a feature of the situation, independently of whether or not this relates to our practical identity: for example, it is just the potential happiness of my daughter that constitutes a reason for me to obey my impulse to buy the toy. But much of the first two chapters of *The Sources of Normativity*, as well as related papers, is spent arguing against realism of this kind.[19] Korsgaard's position here is complex, and has several strands. One strand is that realism is unable to explain the felt obligatoriness of moral reasons,[20] where by relating this to the person's sense of self, we can see why she must act, to preserve her sense of who she is.[21] Another strand is that the realist faces a regress of justification or an arbitrary foundationalism, as either one reason is grounded on another, or the regress is brought to a stop by fiat;[22] by contrast, on her position the reasons that apply to the agent can be explained in terms of her practical identity, where (as we shall see) Korsgaard thinks that the regress can be brought to a satisfactory end. Korsgaard thinks that both of these features mean that her position is better placed that the realist's to deal with the moral sceptic, who asks why she should act morally. A final strand in Korgaard's case against realism is an argument from autonomy: if the reasons we have to act are independent of us, then in acting on those reasons we are not acting freely.[23] On the other hand, if reasons stem from

[19] See (Korsgaard 1996b: 7–89, 2003).

[20] 'According to . . . realism . . . there are facts, which exist independently of the person's mind, about what there is reason to do; rationality consists in conforming one's conduct to those reasons . . . The difficulty with this account in a way exists right on its surface, for the account invites the question why it is necessary to act in accordance with those reasons, and so seems to leave us in need of a reason to be rational . . . we must still explain why the person feels it *necessary* to act on those normative facts, or what it is about *her* that makes them normative *for her*. We must explain how these reasons get a grip on the agent' (Korsgaard 1997: 240).

[21] Cf. (Korsgaard 1996b:100–2).

[22] Cf. (Korsgaard 1996b: 33): 'As these arguments show, realism is a metaphysical position in the exact sense criticized by Kant. We can keep asking why: "Why must I do what is right?"—"Because it is commanded by God"—"But why must I do what is commanded by God?"—and so on, in a way that apparently can go on forever. This is what Kant called a search for the unconditioned—in this case, for something which will bring the reiteration of "but why must I do that?" to an end. The unconditional answer must be one that makes it impossible, unnecessary, or incoherent to ask why again. The realist move is to bring this regress to an end by fiat: he declares that some things are *intrinsically* normative . . . Having discovered that he needs an unconditional answer, the realist straightaway concludes that he has found one.'

[23] (Korsgaard 1996b: 5): 'If the real and the good are no longer one, value must find its way into the world somehow. Form must be imposed on the world of matter. This is the work of

our practical identity, then that makes them intrinsic to who we are, and so compatible with our agency.

Now, of course, all these points against realism can be and have been resisted by realists.[24] But perhaps again we can give Korsgaard the benefit of the doubt here, as many would share her conviction that realism is indeed problematic in the ways that she suggests.

Let us move, then, to the third premise of the argument, where having shown that to have a reason to act this must relate to a particular practical identity, Korsgaard tries to show that no particular practical identity can be adopted unless you adopt the practical identity of being human. I think the idea here is as follows.

Suppose you take an impulse like wanting this toy to be a reason to act because it conforms to your particular practical identity of being a father. But as a reflective agent, you can then ask: what reason have I got to adopt this particular practical identity of being a father? You cannot give as a reason: because then I will go around doing good things like buying toys for my daughter; because doing those things are only reasons for you from the perspective of this identity, which is precisely what is in question. And you cannot give as a reason some further particular practical identity, like being a husband, because the same questions can be raised about that. If you are to halt the regress of reasons, therefore, you must appeal to some reasons that are not grounded in any particular practical identity and so an identity that is likewise not so grounded or 'conditioned': and the only identity of which this is true is the identity of humanity, for without that identity, you could not see yourself as an agent with reasons at all, so it itself does not rest on any reasons given to it by some further particular practical identity. As Korsgaard puts this, beginning with the passage we quoted earlier covering premises 1 and 2:

> It is necessary to have *some* conception of your practical identity, for without it you cannot have reasons to act. We endorse or reject our impulses by determining

art, the work of obligation, and it brings us back to Kant. And this is what we should expect. For it was Kant who completed the revolution, when he said that reason—which is form—isn't in the world, but is something we impose on it. The ethics of autonomy is the only one consistent with the metaphysics of the modern world, and the ethics of autonomy is the ethics of obligation.'

[24] For some notable responses, see (Gaut 1997), (Regan 2002), (Fitzpatrick 2006), (Crisp 2006: 49–56), (Parfit 2006), (Wallace 2006: 71–81). I consider Korsgaard's argument from autonomy further in (Stern 2007a).

whether they are consistent with the ways in which we identify ourselves. Yet most of the self-conceptions which govern us are contingent . . . What is not contingent is that you must be governed by *some* conception of your practical identity. For unless you are committed to some conception of your practical identity, you will lose your grip on yourself as having any reason to do one thing rather than another—and with it, your grip on yourself as having any reason to live and act at all. But this reason for conforming to your particular practical identities is not a reason that *springs from* one of those particular practical identities. It is a reason that springs from your humanity itself, from your identity simply as a *human being*, a reflective animal who needs reasons to act and to live. (Korsgaard 1996b: 121)

I think we can view Korsgaard's position here this way. My practical identity is my sense of who I am. As a reflective agent, I can see that I could be brought to give up any particular practical identity I may have, such as being a father, husband, Englishman, university lecturer, etc., as I come to see that they are not really essential to me as such, in a way that 'alienates' me from them. But I cannot give up my sense of being a person who can think about who he is in the same way, because to do this I would have to be thinking about who I am—and it is this which Korsgaard thinks is distinctive of the practical identity of humanity. In this case, therefore, no more basic identity can supply me with a reason to adopt this identity, any more than some more basic logical principle can supply me with a reason to believe the principle of non-contradiction, in which case I am entitled to treat it as just the sort of 'unconditioned' stopping point that is able to bring the regress of reasons to a principled end.

Now let's consider the final premise: you cannot adopt humanity as a practical identity unless you value your humanity. Thus far, Korsgaard has shown that you are required to adopt humanity as your practical identity if you are to adopt any practical identity at all, as this identity brings a halt to the regress. But suppose you didn't value your identity as human—meaning here being a reflective agent—but just saw it merely as a necessary fact about yourself, which nonetheless you felt neutral about, or even rather regretted and despised?[25] In this case, however, humanity would not bring a halt to the regress, because unless you saw humanity as valuable, it could not give any reasons to act in itself or to adopt any other particular practical identity, where unless it does so, you cannot continue to think of

[25] Cf. (Gibbard 1999: 154): 'Why, though, couldn't I think of reflective choice as a burden, only mitigated by some admirable way that people like me handle it? . . . couldn't I still disvalue the sheer state of being a reflective chooser?'

yourself as a rational agent with reasons. It is therefore a necessary condition of having such reasons that you value your humanity: according to Korsgaard, the price of denying this is to see yourself as living in a world in which there are no reasons to act and thus no way to be an agent at all. In a sense, Korsgaard admits, she cannot prevent the sceptic paying the price if he is determined to do so, in a kind of suicidal abandonment of agency;[26] but in another sense we have little choice but to see ourselves as agents, and so (she thinks) to accept the conclusion of this part of her argument, that you must value your humanity.

IV

Let us call this version of Korsgaard's argument the regress of identities argument. I now want to consider an objection to it.[27]

This concerns whether Korsgaard is right to think that if we work merely with our particular practical identities, we will be threatened with a regress. Now Korsgaard admits, of course, that as a matter of psychology some of us may not reflect on our particular practical identities, and so not face the regress in our daily lives; but that, she argues, just shows that we can be insufficiently reflective and doesn't show we ought not to feel the regress.[28] But, why ought we to feel the regress? Korsgaard's idea seems to be that our particular practical identities are contingent, and that we could therefore always be brought to give them up by finding other identities that are more compelling:

> You may cease to think of yourself as a mother or a citizen or a Quaker . . . This can happen in a variety of ways: it is the stuff of drama, and perfectly familiar to us all. Conflicts that arise between identities, if sufficiently pervasive or severe, may force you to give one of them up: loyalty to your country and its cause may turn you against a pacifist religion, or the reverse. Circumstances may cause you to call the practical importance of an identity into question: falling in love with a Montague may make you think that being a Capulet does not matter after all. Rational reflection may bring you to discard a way of thinking of your practical identity as silly or jejune. (Korsgaard 1996b: 120)

[26] Cf. (Korsgaard 1996b: 160–4).

[27] I also have my doubts about the step from 4 to 5,but will leave these aside for now.

[28] Cf. Korsgaard's comments on the member of the Mafia brought into the discussion by G. A. Cohen: see (Korsgaard 1996b: 257–8).

Now Korsgaard is of course correct that this sort of thing can occur, and when it does, in a way that should lead you to question your particular practical identity. But Korsgaard seems to think that she can generalize from this, to show that it can apply to every particular practical identity for every agent. But I think that the kind of self-doubt that Korsgaard sees as pervasive is in fact harder to achieve than she realizes—where I don't just mean psychologically harder (which I agree with her is irrelevant here), but rationally or normatively. I think one can see this by looking at her examples. It is important to the persuasiveness of her examples, that the competing identity is not just different from the one you hold, but can itself supply reasons in its favour: for example, the pacifist can nonetheless come to see some value in fighting for his country. But might there not be particular practical identities which nonetheless seem invulnerable to competing reasons in this way, despite being identities we accept are contingent, in the sense that we can see we might not have had the identity in question? For example, suppose that I am a loving son. I am certainly conscious that I might not have been, had I been raised in a different way or in a different time or place, so it is a contingent identity in that sense. But suppose someone offers me an alternative identity. Unless they can supply me with reasons to think being a loving son is wrong or mistaken in some way, why should I be brought to doubt my identity? But it might seem that the only reasons that would count as reasons to give up that identity are ones that would only do so once that identity has been given up. So, for example, you might say I should become a ruthless city trader, and so stop wasting my time visiting my sick mother and spend it arranging profitable deals instead. But what reason could you give me for taking this seriously as an identity, given that you are asking me to betray everything I hold dear? It seems that you would need to give me some internal grounds for giving up my identity (for example, that my loving regard for my mother is making her life worse and not better), which even in the case of an identity which is contingent may not be forthcoming. Or you would need to appeal to common ground between being a son and the values that leads to, and being a ruthless city trader, just as the values of the pacifist and non-pacifist may be said to coincide at certain points, from which the divergences can be explored. But insofar as this is possible, then we can continue to operate with reasons at the level of converging particular practical identities, rather than moving to the kind of universal and necessary identity that Korsgaard thinks is required. So, if this is right,

Korsgaard has arguably not done enough to show that it is only the identity of humanity that is invulnerable to the regress issue and hence 'unconditioned', and not also certain particular practical identities; and if this is right, her transcendental argument in this form collapses at step 3.[29]

V

I now want to look at a second transcendental argument that can be found in Korsgaard's work, which I think fares better than the first one.

This second argument is modelled on an argument that Korsgaard finds in Kant, and which she outlines as follows:

> [Kant] started from the fact that when we make a choice we must regard its object as good. His point is the one I have been making—that being human we must endorse our impulses before we can act on them. He asked what it is that makes these objects good, and, rejecting one form of realism, he decided that the goodness was not in the objects themselves. Were it not for our desires and inclinations—and for the various physiological, psychological, and social conditions which gave rise to those desires and inclinations—we would not find their objects good. Kant saw that we take things to be important because they are important to us—and he concluded that we must therefore take ourselves to be important. In this way, the value of humanity itself is implicit in every human choice. If complete normative scepticism is to be avoided—if there is such a thing as a reason for action—then humanity, as the source of all reasons and values, must be valued for its own sake. (Korsgaard 1996b: 122)[30]

This argument can be laid out as follows:

1. To rationally choose to φ-ing, you must regard φ-ing as good.
2. You cannot regard φ-ing as good in itself, but can only regard φ-ing as good because it satisfies your needs, desires, inclinations, and so on.
3. You cannot regard your desiring or needing to φ as making it good unless you regard yourself as valuable.
4. Therefore, you must regard yourself as valuable.

[29] For similar sorts of objections, see (Schneewind 1998: 43–8). See also Korsgaard's reply in (Korsgaard 1998).

[30] Korsgaard claims to take her inspiration from Kant's *Groundwork of the Metaphysics of Morals* (1998: 4:427–8), which she discusses further in her paper 'Kant's Formula of Humanity', reprinted in (Korsgaard 1996a: 106–32). I leave aside here whether or not Korsgaard is correct in her reading of Kant; but for some doubts on this score, see (Timmermann 2006).

Consider this example. To rationally choose to eat this piece of chocolate cake, I must think that eating the cake is good in some way. How can I regard it as good? It seems implausible to say that eating the cake is good in itself, of intrinsic value. It also seems implausible to say that it is good just because it satisfies a desire as such: for even if I was bulimic it might do that, but still not be regarded as good. A third suggestion, then, is that it can be seen as good because it is good for me, as satisfying a genuine need or desire of mine. But if I think this is what makes eating the piece of cake good, I must value myself as, otherwise, I could not hold that satisfying me is sufficient to make something good enough for it to be rational for me to desire it; so I must regard myself as valuable. Put conversely: suppose that you thought that you and your life were utterly worthless, pointless, meaningless—that in your eyes, you were valueless. And suppose that you are faced with a piece of cake: on what basis would you choose to eat it? It seems unlikely that there is something intrinsically good about eating it, or that you should do so just because you find yourself with a desire to do so, even while finding your existence valueless. It seems that the only reason to do so would be if you thought eating the cake brought you some genuine benefit—but if you thought your life was worthless, how could you see this as a reason either? Why is bringing benefit to something that in your eyes is so utterly without value a reasonable thing to do?

There are some dangers in this argument, however. One, which Korsgaard considers, is that it might lead to 'self-conceit':[31] that is, I might conclude from this that I am supremely valuable, simply as Bob Stern, which could obviously then get in the way of my ethical treatment of others. But, this worry might be lessened by the thought that while the argument gets me to see that I must find something valuable about me, it need not be anything about me in particular, and perhaps could instead be something about me that is more general—such as my humanity or personhood. However, while Korsgaard says that reflection will indeed lead us in this more general direction, we will need to see how. A second, perhaps related, worry is that this argument has a troubling parallel in the case of Satan, where Satan goes through 1 to 4 above, and concludes that he must regard his devilish nature as valuable. If this argument somehow entitles us to regard our humanity or personhood as valuable, why doesn't

[31] (Korsgaard 1998: 54). Cf. also (Korsgaard 1996b: 249–50).

it entitle Satan to think the same about his nature? This is not the same as self-conceit, because he is not valuing himself as Satan just qua Satan; he is valuing his nature, just as we are valuing ours. Nor does devilishness seem any less central to his nature than humanity is to ours. So it is hard to see how the Satanic parallel can be avoided by the argument as it stands.

Nonetheless, it is possible that something can be built on the central idea of the argument, which I take to be this: as long as we think we can act for reasons based on the value of things, but at the same time reject any realism about that value applying to things independently of us, then we must be treated as the source of value and in a way that makes rational choice possible. We can therefore see Korsgaard's second argument as attempting something along these lines, using her notion of practical identity to perhaps avoid the two problems we have identified with the Kantian argument.

Here, then, is an outline of Korsgaard's second argument:

1. To rationally choose to φ, you must take it that φ-ing is the rational thing to do.
2. Since X in itself gives you no reason to φ, you can take it that φ-ing is the rational thing to do only if you regard your practical identity as making it rational to φ.
3. You cannot regard your practical identity as making φ-ing rational thing to do unless you can see some value in that practical identity.
4. You cannot see any value in any particular practical identity as such, but can regard it as valuable only because of the contribution it makes to giving you reasons and values by which to live.
5. You cannot see having a practical identity as valuable in this way unless you think your having a life containing reasons and values is important.
6. You cannot regard it as important that your life contain reasons and values unless you regard your leading a rationally structured life as valuable.
7. You cannot regard your leading a rationally structured life as valuable unless you value yourself qua rational agent.
8. Therefore, you must value yourself qua rational agent, if you are to make any rational choice.

The first step is now familiar: to act is to do or choose something for a reason. The second step is also now familiar: Korsgaard thinks that we have

reasons to act because of our practical identities, not because acts have reasons attached to them in themselves. Once again, realists might demur,[32] claiming that some actions are rational things to do, because some things have value as such: so, perhaps knowledge is valuable in itself, thereby making it rational to seek it.[33] But as before, let us leave such worries aside and assume with Korsgaard that nothing is objectively rational for us to do.

The third step asks how a practical identity can make something into a reason for an agent: how can the fact that I am a father make it rational for me to buy my daughter this toy? The thought here is that it can only do so if I see value in that identity. Korsgaard stresses this when she writes:

> The conception of one's identity in question here is not a theoretical one, a view about what as a matter of inescapable scientific fact you are. It is better understood as a description under which you value yourself, a description under which you find your life to be worth living and your actions to be worth undertaking. (Korsgaard 1996b: 101)

So, being a father, whether contingently or essentially, gives one no reason to be a caring or devoted father of a sort that would have good reason to buy a daughter a gift; rather, valuing one's fatherhood does this.

But (moving on to step 4), how can I see my particular practical identity as valuable? I think Korsgaard's position here is that I cannot see any value in any particular practical identity as such, and this might seem to rest on something like the sort of regress argument we looked at and criticized above: one will always be faced with requiring a reason for valuing any particular practical identity, and this regress cannot be halted by any such identity. But, as we saw, this thought can perhaps be resisted. However, I think that even without a regress argument, Korsgaard can make her point here, more by using the objections to realism considered previously: namely, that to see value in any particular practical identity as such is to be committed to realism, to thinking that being a father, an Englishman, a university lecturer or whatever matters as such; or (in a way that is in the end equally realist), it matters because of the intrinsically valuable things it leads you to do. But, as we have seen, Korsgaard also takes such realist

[32] See, for example, (Scanlon 1998: 55–72), and (Kerstein 2002: 70–2).
[33] Cf. (Regan 2002: 272).

positions to be problematic, so can perhaps use such arguments here, without appealing to the regress considerations at all.

So suppose we allow that no particular practical identity can be seen to have value in itself; Korsgaard then offers as the only remaining explanation of its value to the agent that has that identity, that such identities have the general capacity of enabling the agent to live a life containing reasons: because I have whatever particular practical identities I do (father, Englishman, university lecturer . . .), I can then find things to be valuable and act rationally accordingly, in a way that gives me unity as a subject. As Korsgaard puts it: 'To be a thing, one thing, a unity, an entity; to be anything at all: in the metaphysical sense, that is what it means to have integrity. But we use the term for someone who lives up to his own standards. And that is because we think that living up to them is what makes him one, so what makes him a person at all' (Korsgaard 1996b:102).

But then (step 5), to think that this makes having some sort of particular practical identity important, you must think that it matters that your life have the sort of rational structure that having such identities provides; but (step 6), to see that as mattering, you must see value in your leading a rationally structured life. And then, finally, to see value in your leading such a life, you must see your rational nature as valuable, which is to value your humanity.

Does this Korsgaardian argument avoid the pitfalls of the Kantian one discussed earlier? I think it avoids the problem of self-conceit, because it does seem that what you end up valuing is not yourself simply as such, but yourself qua rational agent. And I think as I have presented it, it avoids the problem of the Satanic parallel, because all it shows is that Satan must value his rational nature, not his devilishness.

For both these problems to be avoided, however, it is important to run the argument as I have done, not as it is sometimes presented by Korsgaard, which is via the notion of need.[34] This would follow the same premises as before for 1–5, and then go as follows:

[34] Cf. (Korsgaard 1996b: 121): 'Most of the time, our reasons for action spring from our more contingent and local identities. But part of the normative force of those reasons spring from the value we place on ourselves as human beings who need such identities. In this way all values depends on the value of humanity; other forms of practical identity matter in part because humanity requires them'; and (Korsgaard 1996b: 125): 'Our other practical identities depend for their normativity on the normativity of our human identity—on our own endorsement of our human need to be governed by such identities—and cannot withstand reflective scrutiny without it. We must value ourselves as human.'

6*. You cannot regard it as important that your life contain reasons and values unless you take your need to lead this sort of life as important.
7*. You cannot take this need to be important unless you take yourself to be valuable.
8*. Therefore, you must value yourself, if you are to make any rational choice.

The difficulty with 6*–8*, I think, is that 8* does not stipulate what it is about yourself that you are required to value, so that this could be my sheer particularity (self-conceit), or if I am not in fact human, my non-human nature (Satan). This is because 6* just identifies a need, and says that this need could not be important unless the agent who has the need were seen to be valuable somehow—whereas the previous argument narrows value down to rational agency, and so rules out both self-conceit and devilishness.

VI

I have therefore reconstructed that part of Korsgaard's strategy which offers an argument to the effect that you must value your humanity, as a transcendental argument. It turns out that if it is to be made plausible in this way, a lot depends on accepting Korsgaard's arguments against realism; but then, many have suspected that some commitment to anti-realism is required to make a transcendental argument convincing. A worry then is that it can appear to make the argument redundant in the standard anti-sceptical case, because anti-realism appears sufficient as a response to scepticism on its own;[35] but in this ethical case, this does not seem to be the issue, so that here this worry is less of a concern. Of course, as Korsgaard herself allows, this transcendental argument in itself is not meant to be sufficient to complete her project, which still requires a third phase, which I have not considered, and which may still be found to be problematic.[36] Nonetheless, I hope to have shown something, which

[35] For more discussion of this issue, see (Stern 2000: 49–58).

[36] There is another objection which I also cannot address in this chapter, and which is not considered directly by Korsgaard: namely that if 'valuing your humanity' comes down to 'valuing your rational agency', will that cover enough of what morality is usually meant to cover? Allan Gibbard puts the point this way: 'If valuing my humanity is taking pride in being a reflective chooser, how does that constrain what I do, what I reflectively choose? Perhaps I

is how the transcendental argument of the second phase can be seen to work, and how it is more plausible than many of Korsgaard's critics have found it; and this I think is an achievement of sorts.[37]

References

Bittner, R. 1989. *What Reason Demands*. Cambridge: Cambridge University Press.

Cohen, G. A. 1996. Reason, Humanity, and the Moral Law. In Korsgaard 1996b: 167–88.

Crisp, R. 2006. *Reasons and the Good*. Oxford: Oxford University Press.

Fitzpatrick, W. J. 2006. The Practical Turn in Ethical Theory: Korsgaard's Constructivism, Realism, and the Nature of Normativity. *Ethics* 115: 651–91.

Gaut, B. 1997. The Structure of Practical Reason. In G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*: 161–88. Oxford: Oxford University Press.

Gibbard, A. 1999. Morality as Consistency in Living: Korsgaard's Kantian Lectures. *Ethics* 110: 140–64.

Ginsborg, H. 1998. Korsgaard on Choosing Nonmoral Ends. *Ethics* 109: 5–21.

Hume, D. 1965. Of Suicide. In A. MacIntyre (ed.), *Hume's Ethical Writings*: 297–306. New York: Collier Books.

Illies, C. 2003. *The Grounds of Ethical Judgement*. Oxford: Oxford University Press.

Kant, I. 1998. *The Groundwork of the Metaphysics of Morals*. Translated and edited by M. Gregor. Cambridge: Cambridge University Press.

Kerstein, S. J. 2002. *Kant's Search for the Supreme Principle of Morality*. Cambridge: Cambridge University Press.

Korsgaard, C. M. 1996a. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.

—— 1996b. *The Sources of Normativity*. Cambridge: Cambridge University Press.

must nurture my powers of reflective agency . . . Still, if that's all we must do as human beings, then enlightenment morality is far too narrow: we'll need to oppose pain and seek fulfilment and enjoyment only when they affect our powers of reflective agency' (Gibbard 1999: 156). I am not sure the Kantian would in fact see much of a worry here. Perhaps a deeper concern, however, is that even if the argument I have considered works, there is a problem with Korsgaard's whole *strategy* of attempting to respond to moral scepticism in the way she does, as in the end the reason to be moral becomes grounded in the interest we have in being agents, which is to distort the nature of genuine moral action which should not be grounded in anything *outside* morality itself. I discuss this issue further in (Stern 2010).

[37] I gave a first version of this chapter at one of the Transcendental Philosophy and Naturalism workshops, under the AHRC funded project run by Mark Sacks. Mark offered very useful and characteristically generous comments on that occasion, the last on which I was to see him. I am therefore particularly grateful to Joel Smith and Peter Sullivan for undertaking the publication of this chapter, as what is very sadly the last chapter in the discussion that Mark and I had on matters transcendental over many years, from which I learned so much.

—— 1997. The Normativity of Instrumental Reason. In G. Cullity and B. Gaut (eds.), *Ethics and Practical Reason*: 215–54. Oxford: Oxford University Press. Reprinted in Korsgaard 2008: 27–68.

—— 1998. Motivation, Metaphysics, and the Value of the Self: A Reply to Ginsborg, Guyer, and Schneewind. *Ethics* 109: 49–66.

—— 1999. Self-Constitution in the Ethics of Plato and Kant. *Journal of Ethics* 3: 1–29.

—— 2003. Realism and Constructivism in Twentieth-Century Moral Philosophy. *APA Centennial Supplement to the Journal of Philosophical Research*: 99–122. Reprinted in Korsgaard 2008: 302–26.

—— 2008. *The Constitution of Agency: Essays on Practical Reason and Moral Psychology*. Oxford: Oxford University Press.

—— 2009. *Self-Constitution*. Oxford: Oxford University Press.

Nagel, T. 1996. Universality and the Reflective Self. In Korsgaard 1996b: 200–9.

Parfit, D. 2006. Normativity. In R. Shafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. I: 325–80. Oxford: Oxford University Press.

Regan, D. H. 2002. The Value of Rational Nature. *Ethics* 112: 267–91.

Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.

Schneewind, J. B. 1998. Korsgaard and the Unconditional in Morality. *Ethics* 109: 36–48.

Skidmore, J. 2002. Skepticism about Practical Reason: Transcendental Arguments and their Limits. *Philosophical Studies* 109: 121–41.

Skorupski, J. 1998. Rescuing Moral Obligation. *European Journal of Philosophy* 6: 335–55.

Stern, R. 2000. *Transcendental Arguments and Scepticism*. Oxford: Oxford University Press.

—— 2007a. Freedom, Self-Legislation and Morality in Kant and Hegel: Constructivist vs Realist Accounts. In E. Hammer (ed.), *German Idealism: Contemporary Perspectives*: 245–66. London: Routledge.

—— 2007b. Transcendental Arguments: A Plea for Modesty', *Grazer Philosophische Studien* 74: 143–61.

—— 2010. Moral Scepticism and Agency: Kant and Korsgaard. *Ratio* 23: 453–74.

Street, S. 2008. Constructivism about Reasons. In R. Schafer-Landau (ed.), *Oxford Studies in Metaethics*, vol. III: 207–46. Oxford: Oxford University Press.

Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241–56.

Timmermann, J. 2006. Value Without Regress: Kant's 'Formula of Humanity' Revisited. *European Journal of Philosophy* 14: 69–93.

Wallace, R. J. 2006. Normativity and the Will. Reprinted in his *Normativity and the Will*: 71–81. Oxford: Oxford University Press.

Wood, A. W. 1998. Review of Korsgaard, *Creating the Kingdom of Ends*. *Philosophical Review* 107: 607–11.

—— 1999. *Kant's Ethical Thought*. Cambridge: Cambridge University Press.

6

Reasons, Naturalism, and Transcendental Philosophy

Hilary Kornblith

There is a view about knowledge, and about the nature of reason, which is widely held among naturalists. Knowledge, on this view, is a natural phenomenon.[1] Human beings, and many other animals, know a great many things, and it is in virtue of this knowledge that they are able to negotiate their environment so successfully. If we want to understand what knowledge is, and how knowledge is possible, then we need to examine this natural phenomenon in the very same way that we examine other natural phenomena: that is, empirically. In order to understand how knowledge is possible, we need to understand certain large-scale features of the mind, as well as certain large-scale features of the environment. When features of the mind dovetail in certain characteristic ways with features of the environment, knowledge becomes possible. Understanding the way in which our minds are structured to pick up information about the environment allows us to appreciate what makes knowledge possible, and to understand how our minds are responsive to reasons. Two brief illustrations will perhaps be useful.

Consider, first, the Chomskian revolution in linguistics. Chomsky argued that language is learnable for humans only in virtue of certain structural features of the mind.[2] The mind, on this view, has a modular

[1] I have defended my own version of this view in two books: (Kornblith 1993, 2002).
[2] For early presentations of this view, see (Chomsky 1957, 1965).

structure. There is a language-learning module, outfitted with certain presuppositions about the structure of humanly learnable first languages. While such languages might, in principle, have had quite a different structure from the one they in fact have, any attempt to discern the structure in spoken language without making quite substantial presuppositions about that very structure would be doomed to failure. This is the poverty of the stimulus argument. Language learning is possible for us only because the mind does not approach the problem without any presuppositions. Once we see how the mind is specially outfitted for the task of language learning, we see how the linguistic data that we make use of in coming to learn the structure of our first language provide us with all the reasons we need to comprehend the structure of our linguistic environment. The Chomskian picture, of course, is not supported by way of *a priori* argument. It is, instead, based on experimental evidence about the early linguistic environment of children, about the tempo and mode of language acquisition, about the structure of human languages themselves, and about the structure of the human mind and the human brain.

Consider, second, work on visual information processing.[3] Here again, we see the modular structure of the mind at work.[4] The visual system has certain presuppositions about the environment built into it, and it is only in virtue of approaching the task of visual information processing with these presuppositions about our environment that we are able to form any beliefs at all about the world around us. These presuppositions built into the visual system are not true of every logically possible environment, but they are largely true of the environment we inhabit. We presuppose, for example, that the world is largely inhabited by three-dimensional objects having relatively stable boundaries, and it is in virtue of making this presupposition that the visual system is able to detect the geometrical structure of objects around us. It is also in virtue of making this presupposition that the visual system is subject to certain illusions, misperceptions of features of the environment on those rare occasions when the presuppositions are false. Our characteristic pattern of errors may thus be used to help reveal the presuppositions of the visual information processing system,

[3] See, for example, (Marr 1982).

[4] The importance of modularity in understanding the large-scale structure of the mind was emphasized by Jerry Fodor (1983).

thereby allowing us to understand how knowledge of the visual environment becomes possible for us. In coming to see how the visual system works, we come to understand how the data we take in visually give us sufficient reason to form the beliefs we do—beliefs which are largely accurate about the environment around us.

While some of the details of these accounts are of interest only to linguists or perceptual psychologists, we may come to appreciate something about knowledge itself, and about how responsiveness to reason is possible, in seeing the general features which are common to these various accounts. On this kind of view, philosophy may ask questions which are more abstract than the special sciences, but the difference between philoso phy and science is merely a matter of degree of abstractness, rather than any real difference in kind. An understanding of these abstract issues about the possibility of knowledge and the nature of reasons can only be achieved by way of empirical investigation.

There is, however, another way to view knowledge and the nature of reason, a way of looking at these matters that has its origin in Kant. We see this view in quite a number of contemporary philosophers, many of whom make plain their indebtedness to Kant, and many others on whom the Kantian influence is largely mediated by way of Wilfrid Sellars. Thus, for example, Michael Williams tells us that, 'Knowledge is not a natural phenomenon' (Williams 2004: 194). And John Haugeland, endorsing much the same view, tells us that the capacity for knowledge is 'not just a biological or "natural" capacity' (Haugeland 1998: 2). These authors tend to see a very large difference between human and animal cognition, arguing not only that non-human animals are incapable of genuine knowledge, but, more than this, that they are incapable of having genuine beliefs.[5] Many of these authors stress, in a Kantian spirit, the importance of our ability to reflect on our beliefs and our reasons for them. Our ability to reflect on our own mental states not only sets us apart from other animals, but it reveals the decidedly anti-naturalistic character of this approach. Thus, Richard Moran, in discussing the role that self-knowledge plays here, remarks that 'a non-empirical or transcendental relation to the self is ineliminable' (Moran 2001: 90). This special relation to the self, and its role in understanding knowledge and the nature of reasons, comes to the fore,

[5] On this issue, see also, of course, Donald Davidson, especially (Davidson 1984, 2001).

on this view, when we recognize the importance of epistemic agency. As Christine Korsgaard comments, 'It is because of the reflective character of the mind that we must act, as Kant put it, under the idea of freedom' (Korsgaard 1996: 94).

There are, of course, many differences among these authors, but there is, as well, a great deal that they have in common. There is a constellation of issues which these authors see as deeply connected, involving reflection, epistemic agency, normativity, and, in the end, an anti-naturalistic account of knowledge and the nature of reasons. In this chapter, I attempt to give a sympathetic presentation of this anti-naturalistic picture, but I also want to explain why it is that I reject it.[6]

I

Let me begin, then, by sketching an anti-naturalistic picture of some of the distinctive features of human cognition.

Human beings differ from other animals in a number of fundamental ways. First, human beings are language users, while other animals are not. And second, leaving aside, for a moment, the question of whether non-human animals have genuine beliefs, it is quite clear that non-human animals are incapable of reflecting on their mental states and forming beliefs about them. Thus, human beings not only form beliefs about the world around them; they form beliefs about their own beliefs. Non-human animals cannot do this. This ability to reflect on our own mental states marks a crucial difference between humans and other animals.

What is so important about this ability to reflect on one's own mental states? Let us suppose, just for the sake of argument, and just for the moment, that non-human animals do have genuine beliefs. Then the difference we are discussing amounts to this: we humans have both first-order beliefs—beliefs about the world around us, for example—and second-order beliefs as well—beliefs about our first-order beliefs; non-human animals have only first-order beliefs. In the case of non-human animals, their first-order beliefs are produced by a variety of different mechanisms, and many of these mechanisms, no doubt, reliably deliver accurate information about the

[6] I have broached these issues before in two other papers: (Kornblith 2007) and 'The Myth of Epistemic Agency' (Manuscript).

world around them, allowing these animals successfully to negotiate their environments. Of course, these belief-producing mechanisms are not perfectly reliable; they sometimes deliver false beliefs. And more than this, it may well be that some of these mechanisms are not reliable at all; they deliver false beliefs more often, perhaps far more often, than true ones.

Now we humans, of course, have mechanisms that generate first-order beliefs in just this way, and, just as in the case of other animals, many of these mechanisms reliably produce true beliefs, even if they do not do so infallibly. And probably many of these mechanisms are not only less than perfectly reliable; some of them are, in fact, simply unreliable. But while non-human animals are bound, once and for all, to form beliefs by way of these hard-wired mechanisms, human beings are not. Unlike other animals, we are capable of reflecting on our beliefs and the ways in which they are produced, and when we find that our beliefs fail to meet our standards, we are able to intervene in the belief-producing process, making changes in the ways in which we come to form our beliefs. It is thus the ability to reflect on our own mental states which allows for the possibility of changes in the processes by which our beliefs are formed.

This ability to reflect on our own beliefs, and thereby to engage in epistemic self-assessment, thereby makes room for epistemic agency. While other animals are both the unknowing beneficiaries and the unwitting victims of their native processes of belief acquisition, we are active parties in the business of cognition. Non-human animal cognition is simply passive, but, in the human case, we are able to take an active role in moulding our cognitive processes in our own image. It is for this reason that we may sensibly speak of human beings as epistemically responsible agents, and it is in virtue of this that we may hold humans responsible for the beliefs they hold. It is only because we are epistemically responsible that it makes sense to evaluate beliefs as justified or unjustified. While non-human animals may form beliefs which are true or false, the animals themselves play no role in determining how their beliefs are formed, and thus they may not be held responsible for the success or failures of their cognitive processes. Their belief-forming processes may work well or badly, but it would simply be a mistake to think of the resulting beliefs as therefore ones which the animals are either justified or unjustified in holding. There can be no justification in animals who are not responsible for their beliefs. Only human beings may hold their beliefs justifiedly or unjustifiedly.

Since knowledge plausibly requires at least justified, true belief, we see that, strictly speaking, only human beings are capable of genuine knowledge. It is true, of course, that we often speak, informally, of non-human animals as knowing various things, but we also speak—obviously metaphorically—of various mechanical objects, which clearly do not even have beliefs, as knowing various things: the automatic door-opener 'knows' when someone is approaching the door; the GPS system in your car 'knows' when you have made a wrong turn, and so on. Genuine knowledge, however, is the exclusive property of human beings.

Indeed, once we see just how different human cognition is from what goes on in other animals, it starts to become clear that it is not merely talk of animal knowledge which must be seen as metaphorical. Talk even of animal belief cannot be understood as literally correct. Animals certainly have information-bearing states which often accurately reflect features of the world around them, and it is for this reason that the metaphor of 'animal belief' comes so trippingly off the tongue. But we should note that there is a very large difference between having genuine beliefs and merely possessing information-bearing states, even information-bearing states which are implicated in the control of behaviour. Thus, consider the phenomenon of phototropy: plants are sensitive to the presence of sunlight. Internal states of plants reliably register the presence of light, and these internal states are instrumental in moving the plant's leaves in ways so as to expose a larger surface area to the rays of the sun. While these internal states of the plant reliably register information about the presence of sunlight, no one should regard—and almost no one does regard—these internal information-bearing states as beliefs. So the mere fact that non-human animals have internal, information-bearing states which are implicated in the production of behaviour cannot, by itself, give us reason to think that non-human animals have genuine beliefs.

In the case of plants, their information-bearing states may bring about certain motions, but these motions do not constitute actions on the part of the plant. The states of the plant which register the presence of sunlight are part of the 'space of causes', as Sellarsians like to put it; they are not part of the 'space of reasons'. But now we see that the same is true of the information-bearing states of non-human animals: they too are causes of bodily motions, but they do not provide reasons for animals to act or to believe. In order for a state to provide a reason for action, or for a state to provide a reason for belief, it must be one which the agent is capable

of regarding as a reason. Without the ability to reflect on their own mental states, non-human animals do not have this ability. They do not have the concept of a reason, or of a belief, and without these concepts, they therefore cannot have either reasons or beliefs. Their internal information-bearing states form a network of mere causes. It is only in creatures capable of reflection, capable of standing back from their own first-order processes of belief acquisition, capable of self-assessment, and capable of epistemic agency, that internal information-bearing states may serve as reasons for belief or action, and thus that these states may be regarded as genuine beliefs. Human beings are thus not only the only genuine knowers; we are the only creatures with genuine propositional attitudes of any kind.

On this view, indeed, not all human beings are genuine knowers or even genuine believers. Neonates are not, nor are young children. They, like non-human animals, do not have the concept of belief, or of reason, or of truth, all of which are needed to have the kinds of second-order mental states needed to be epistemically responsible agents. But the relevant concepts are surely had by all normal human adults, and although we do not, of course, self-consciously review every act of belief acquisition to see that it measures up to our standards, nothing like this is necessary in order to be epistemically responsible. We do, on occasion, stop to ask ourselves whether a certain belief is really one which we should endorse, and, in doing so, we thereby exercise our epistemic responsibility. On some occasions, we come to the conclusion that a belief we already hold, or one we are naturally inclined to accept, is indeed worthy of our acceptance. But we also sometimes discover that such beliefs are not worthy. In such cases, we engage in acts of belief revision, or we resist a temptation to believe. More than this, we may undertake self-conscious acts designed to revise the ways in which we form beliefs in the future. We resolve to be more careful, for example, in evaluating evidence; we resolve not to be taken in by a well-spoken or attractive speaker without more carefully examining the reasons offered; we resolve to think things through rather than jump to conclusions, and so on.

As some authors emphasize, we should not think that the exercise of our epistemic responsibility is entirely, or even primarily, a private affair. We not only engage with reason when we reflect on the legitimacy of our beliefs, but in dialogue with others when we give and ask for reasons. We ask, and are asked, why it is that some particular belief is held, and in doing so, we hold each other responsible for our beliefs. Non-human animals,

and very young children, do not engage in the activity of giving and asking for reasons, since they do not speak a language. This important dimension of epistemic responsibility is thus not only essentially social, but essentially linguistic.

Between private acts of deliberation and evaluation, and social acts of giving and asking for reasons, adult human beings are able to take charge of their cognitive faculties. Information processing in lower creatures is simply given for them by nature, but genuine cognition involves the kind of self-constitution which can only be found in epistemically responsible agents. Much of the way in which adult human beings come to take on new beliefs is shaped by their epistemic activity and is not simply determined by their biological natures. But even those processes of belief acquisition which survive both private reflection and public review should not be viewed, in adult humans, as merely due to the providence of nature. The very fact that it has survived such review makes us responsible for this part of our cognitive doings, just as much as the part which we actively initiate ourselves.

Cognition in adult human beings is thus far more complicated than the kind of information processing which may be found in non-human animals as well as young children. Human adult cognition crucially involves self-conscious deliberation, self-evaluation, the social practice of giving and asking for reasons, and thus genuine epistemic agency and epistemic responsibility, none of which is found in young children or other animals. Once we recognize this, we see that it is only in the case of human adults that there is any kind of genuine engagement with reason, rather than the simple workings of some sort of causal mechanism. And it is for this reason that human cognition must be seen, not merely as something more complicated than the information processing that goes on in lower animals, but as something which cannot be fully captured by a naturalistic world view.

It is this move, of course, which is crucial: the move from the space of causes—which is occupied by the information processing mechanisms present in non-human animals and young children—to the space of reasons—which is occupied only by adult human beings—is a move that shows the limitations of the naturalistic picture. Naturalists regard human cognition as just one more natural phenomenon, no different in kind from the information processing that goes on in other animals, or, for that matter, from the interaction between salt and water when the two are

mixed and sodium and chlorine ions go into solution. From a naturalistic perspective, not only are chemical interactions to be regarded as revealing the operation of natural laws, but the same is true of the information processing that goes on in animals, and, most importantly, so too is the operation of human reason. Human reason is seen, on the naturalistic picture, as yet one more natural phenomenon among many, bound, like chemical interactions, by the operation of causal laws. But what the story I have been telling here is meant to reveal is just how misguided the naturalistic picture is. When information processing is reduced to the mechanical operation of the space of causes, we inevitably leave out what is distinctive of human cognition: our ability to take control of our cognitive lives through the exercise of our epistemic agency as revealed in the activity of reflecting on our own cognitive states and processes. It is only because we are epistemically responsible agents that we may be properly understood as engaging with reason at all. To put the point just a bit differently, but perhaps a bit more familiarly, the naturalistic picture of human beings, in giving a thoroughly descriptive account of the nature of human cognition, inevitably leaves normativity out of the picture. But from the perspective of someone trying to understand what human reason is all about, there is a certain irony to be found here, for the naturalists thereby succeed in providing the kind of account they aspire to only at the cost of eliminating the very phenomenon—the workings of reason—that they seek to accommodate.

II

I hope that the anti-naturalistic picture I have just sketched is neither unfamiliar nor entirely unappealing. Its main elements draw on certain Kantian views about the nature of reasons, and, at the same time, on certain highly commonsensical views about the differences between human cognition and non-human information processing. Here, I would like to highlight what I see as the central claims of the view, and I would also like to make clear (in the footnotes) that I am not attacking a straw man of my own invention. We see important elements of this view in the work of Robert Brandom, Donald Davidson, John Haugeland, Christine Korsgaard, John McDowell, Richard Moran, and Michael Williams.

1. Adult human beings differ from non-human animals and young children in that the former, but not the latter, have the ability to reflect on their mental states and thus the ability to form second-order beliefs.[7]
2. The mechanisms by which adult human beings form their beliefs change over time with the acquisition of new information, while the mechanisms of information processing in non-human animals and young children are fixed.[8]
3. Points 1 and 2 above are not unrelated. Adult human belief-producing mechanisms may change over time only because we have the ability to reflect on our beliefs and the ways in which we arrived at them. When we find that we have arrived at our beliefs in ways which we do not endorse, we are able to modify the ways in which we subsequently form our beliefs.[9]
4. The mechanisms of belief acquisition may thus change over time only in creatures who have the concept of belief, the concept of reason, and the concept of truth.[10]

[7] This point is taken for granted by all of the authors listed above. It is certainly a common-sense view, and there is, as well, a good bit of support for it in the cognitive ethology literature, although none of the authors discussed here refers to that literature. Relevant work in ethology includes (Povinelli and Eddy 1996), (Tomasello and Call 1997), and (Povinelli 2000).

[8] Thus, for example, Michael Williams remarks, 'We might say, animals don't need the capacity for epistemic assessment because they don't test hypotheses: they test themselves. But this is why they are not truly sensitive to reasons. They cannot really change their minds, though the information-acquiring and processing capacities of the species can change over time' (Williams 2004: 207). Williams surely implies here that, while 'the information-acquiring and processing capacities of the species can change over time', they do not change within the life of an individual animal.

[9] See the quote from Williams in note 8 above. But also see (Brandom 1994: 199–271, 2000: 97–122), (Korsgaard 1996: 90–130), and (McDowell 1996: 108–26). Williams remarks, 'The Sellarsian account of observation, which I share with Brandom, does not require constant self-monitoring. But it does require the capacity to allow for observational errors, and a capacity to rethink the significance of putative observational evidence, should the need arise (e.g., if we find reason to think that the conditions of observation are not standard). Thus, the responsible handling of observational beliefs requires not just the concepts of truth and reliability, but an extensive knowledge of our observational capacities, particularly the errors to which they are subject' (Williams 2004: 208). And McDowell comments, 'We cannot construe [mere animals] as continually reshaping a world-view in rational response to the deliverances of experience; not if the idea of rational response requires subjects who are in charge of their thinking, standing ready to reassess what is a reason for what, and to change their responsive propensities accordingly' (McDowell 1996: 114).

[10] This is, of course, a central theme in Davidson's work (see especially the work cited in note 5 above). See also the work by Brandom, McDowell, and Williams cited in note 9. Thus, for example, Williams remarks on the difference between those creatures who have beliefs and those who do not: 'Believers recognize that their beliefs are sometimes false, and

5. Only adult human beings may thus be properly regarded as epistemically responsible. We alone are epistemic agents, active in the ways in which we arrive at our beliefs, rather than merely passive information processors.[11]
6. Thus, only adult human belief is apt for normative assessment. Our beliefs alone may be justified or unjustified. Our beliefs alone may be responsive to reason.[12]
7. Thus, only adult human beings are capable of genuine knowledge.[13]
8. Because animal information processing is merely a matter of the workings of certain causal mechanisms operating within the animals, rather than the kind of responsiveness to reason one sees in adult human beings, it is a mistake even to regard non-human animals as having beliefs.[14]
9. Thus, while the kind of information processing which takes place in non-human animals may be fully understood by way of the sciences, adult human cognition is not a natural phenomenon.[15]

III

Let me now turn to critical discussion of these claims.

Claim 1—that only adult human beings have second-order beliefs—is common ground to all the authors under discussion. Ironically, although the literature in cognitive ethology strongly supported this view until recently, there is now some reason to be more cautious in endorsing this claim.[16] But let us allow this point to stand for the sake of argument. Every last one of the remaining claims, however, should be rejected.

then they change their minds. This is what it is to be sensitive to reasons, thus to be a believer rather than a functionally characterized information processor' (Williams 2004: 207).

[11] See the work by Brandom, Korsgaard, McDowell, and Williams cited in note 9. This is also a central theme of Richard Moran's (2001).

[12] See, again, the works cited in note 9.

[13] See, again, the works cited in note 9.

[14] In addition to the works cited in note 9, and the works of Davidson cited in note 5, see (Haugeland 1998).

[15] See the works cited by Brandom, Haugeland, Korsgaard, Moran, and Williams. McDowell insists that he is not rejecting naturalism, but only an extreme form of it.

[16] See, for example, (Tomasello 2003).

Thus, let us turn to the second claim—that the mechanisms of information processing in non-human animals are fixed once and for all, unlike in the human case. This claim is demonstrably false. There are, of course, some mechanisms like this in non-human animals. Famously, frogs are responsive to small moving objects very close to their faces (Lettvin et al. 1965). When flies come within close range, the frog's tongue lashes out, grabbing hold of the fly, and the fly is then swallowed. Similarly, if a BB is rolled close to the frog's face, it will grab hold of the BB, and swallow it just as well. While there is nothing at all noteworthy about this behaviour, since we all make mistakes, it is worth noting that the frog will do this again if a second BB is rolled its way; and it will do it a third time; and a fourth, and so on. The frog simply doesn't learn from its experience with the BBs. Similarly, as Dean Wooldridge notes,

> When the time comes for egg laying, the wasp *Sphex* builds a burrow for the purpose and seeks out a cricket which she stings in such a way as to paralyze but not kill it. She drags the cricket into the burrow, lays her eggs alongside, closes the burrow, then flies away, never to return. In due course, the eggs hatch and the wasp grubs feed off the paralyzed cricket, which has not decayed, having been kept in the wasp equivalent of deep freeze. To the human mind, such an elaborately organized and seemingly purposeful routine conveys a convincing flavor of logic and thoughtfulness—until more details are examined. For example, the wasp's routine is to bring the paralyzed cricket to the burrow, leave it on the threshold, go inside to see that all is well, emerge, and then drag the cricket in. If the cricket is moved a few inches away while the wasp is inside making her preliminary inspection, the wasp, on emerging from the burrow, will bring the cricket back to the threshold, but not inside, and will then repeat the preparatory procedure of entering the burrow to see that everything is all right. If again the cricket is removed a few inches while the wasp is inside, once again she will move the cricket up to the threshold and re-enter the burrow for a final check. The wasp never thinks of pulling the cricket straight in. On one occasion this procedure was repeated forty times, always with the same result. (Wooldridge 1963: 82. Quoted in (Dennett 1984))

The second claim amounts to the suggestion that information processing in non-human animals is always like this. This amounts to the suggestion that non-human animals do not learn from their experience.

We need not look to work with primates to see that this is not even close to the truth. Skinner's early work with rats and pigeons shows that they are responsive to changes in their environment, and that they easily learn how to go about attaining a variety of rewards, displaying a subtle

sensitivity to the ways in which their environment has changed (Skinner and Ferster 1957). The mechanisms of learning in animals are complex and varied, as even the most conservative writers on this topic allow.[17] The simple reflexes we see in the case of the frog and the wasp are not the rule in animal information processing; they are not even the rule in frogs (Kelley 2004) or wasps (Evans 1966). Even in the case of fairly stereotyped behaviours—such as the broken-wing display in piping plovers, used to mislead would-be predators—new information may often be integrated with old in ways completely unlike the hard-wired fly-swallowing behaviour of frogs and the striking perseveration of *Sphex* (Ristau 1991). The fact of animal learning has been well documented for as long as there has been serious work on animal behaviour. Animal learning requires the integration of new information with old, and this, in turn, makes itself manifest in the ways in which still later information is processed. All of this is quite prosaic. More interesting, and far more complex, are the phenomena of problem solving and innovation.[18] But we need not examine such subtle phenomena in order to see that the manner in which animals process information may change over time.

All of this puts the lie, of course, to the third claim as well: that the manner in which adult human beings process information may change over time only as a result of our ability to reflect on our own mental states. Human beings, do, of course, sometimes reflect on their beliefs and the manner in which they came about. We do, as a result of such reflection, sometimes change the ways in which we subsequently reason. But just as non-human animals integrate new information with old without the need for second-order beliefs, human beings do the very same thing. To take a single example: when a student makes an appointment to talk with me in my office, I come to form the belief that the student will arrive in my office at roughly the appointed time. But if a particular student makes such an appointment and then fails to keep it, and then does the very same thing again, I no longer form the belief that this particular student will arrive in my office simply because he tells me that he will. When my expectations are defeated, I come to respond quite differently to being told that the student will show up for his appointment. This change in the way I respond—in particular, this change in the inferences I draw—is not

[17] For one conservative survey, see (Shettleworth 1998: 95–232).

[18] See, for example, (Reader and Laland 2003).

typically accomplished by reflecting on my earlier mistake. I don't need to reflect on such things in order to stop drawing the conclusion that the student will be in my office. Instead, my inferential behaviour simply changes in response to the information that the student has regularly failed to show up for appointments he has scheduled. To suggest that a higher-level belief is required here in order for any inferential change to occur is to make the very mistake that Lewis Carroll (1895) so picturesquely warned us against.[19]

It is also worth pointing out here that we surely tend to over-estimate the extent to which reflection serves as a driving force for epistemic change. It is not merely that epistemic change does not require the intervention of reflection on our beliefs and the manner in which they were acquired. Even when we do reflect on our first-order mechanisms of belief acquisition and retention, the manner in which our beliefs are in fact acquired is not transparent to introspection. The beliefs we form about our own mechanisms of belief acquisition are often inaccurate, and our attempts to monitor the ways in which we form our beliefs in order to improve the accuracy with which we reason very frequently results in greater self-confidence, but no greater reliability.[20] Introspection provides us with the illusion that we have an extremely active role to play in monitoring and controlling the ways in which we form beliefs, but, in fact, our reflective activity is often epiphenomenal with respect to the ways in which we reason.[21]

This bears, as well, on Claim 4, that the mechanisms of belief acquisition may change over time only in creatures who have the concepts of belief, of reason, and of truth. Notice just how radical this claim is. It is taken for granted by the writers under discussion here that such sophisticated concepts are not possessed by non-human animals, nor are they possessed by young children. This is not unreasonable.[22] Indeed, the standard story about cognitive development in children has it that early on they have beliefs about the world around them; only much later are they able to

[19] I have discussed this problem further in section IV of (Kornblith 2007) and in section II of (Kornblith 2010).

[20] I have argued for these points at length in Chapter 4 of (Kornblith 2002).

[21] Again, see (Kornbith 2002), Chapter 4, for details.

[22] But see some reason for caution about mental state concepts in the work cited in note 16. No one that I know of has suggested, however, that non-human animals or young children have concepts of reason or of truth.

form beliefs about mental states, let alone about reasons qua reasons, or about truth.[23] But if we acknowledge the conceptual limitations of non-human animals and young children while simultaneously insisting that great conceptual sophistication is required in order for the mechanisms of belief acquisition to change over time, then one is left with an extremely unpalatable dilemma: either insist that, outside of adult humans, the mechanisms of belief acquisition never change—that is, that learning never takes place—which, as we have seen, is manifestly false; or, alternatively, adopt the still more radical view that there simply are no beliefs in any creatures other than adult humans. The writers under discussion adopt this more radical view. We will discuss this issue directly when we turn to Claim 8.

So let us turn to Claim 5, the suggestion that only adult human beings are epistemically responsible agents because only we are active with respect to our belief acquisition, while the information processing which takes place in children and other animals is entirely passive.[24] The picture we are supposed to endorse here is that first-order information processing—the sort that goes on in the absence of reflection on one's own mental states, their origin, and the relationship of their contents to one another—is merely passive, and it is passive precisely because it is nothing more than a causal process that goes on within these limited creatures, rather than something that these creatures actually do. But if the mere fact that first-order information processing is a causal process—or perhaps the fact that it is subsumable under causal laws, as some writers insist[25]—thereby makes it something passive, something that merely takes place in these creatures rather than something they do, then what exactly are we supposed to believe about the process of reflection in adult human beings? Is the process of reflection somehow acausal? Is it supposed to be something which somehow eludes causal laws? It is very hard to see how this could be so. Should we really believe that first-order information processing is firmly embedded in the causal structure of the physical world, but reflection on one's mental states somehow takes place outside that causal structure? And where, precisely, is that? This would require that we endorse an extremely radical metaphysical view. The motivation for such a view is not

[23] See, for example, (Astington, Harris, and Olson 1988) and (Gopnik and Meltzoff 1997).

[24] I have discussed this issue in detail in 'The Myth of Epistemic Agency.' I summarize the results from that paper here.

[25] See, for example, (McDowell 1996: Lecture VI).

aided, of course, by the fact that psychologists have found the processes involved in reflection to be just as susceptible to empirical investigation as information processing involving first-order states.[26] Those who actually investigate the workings of reflection find that it is no different in kind, no less embedded in the causal structure of the world, than first-order information processing. Richard Moran's suggestion that there is something 'non-empirical or transcendental' here, that second-order information processing has some special feature that makes it altogether different from, and less susceptible to naturalization than first-order processing, flies in the face of our best available theories. Reflection is no more or less active than first-order belief acquisition.

It is worth pointing out that the suggestion that the reflective/unreflective distinction tracks the active/passive distinction is not even *prima facie* plausible. While reflection is, at times, actively initiated, it is clearly something which we may simply find ourselves engaged in, something which thus goes on in us passively. And if reflection counts as active on those occasions when it is initiated as a result of voluntary activity, then the fact that first-order processes of visual scanning are often voluntarily initiated—for example, when we choose to turn our heads to look at something—should thereby make certain first-order processes of belief acquisition count as active as well.

This casts doubt, as well, on Claim 6, that only adult human belief is apt for normative assessment; only our beliefs may be responsive to reason. We should not think that responsiveness to reason requires beliefs about reasons. If in order to be responsive to A as a reason for believing B, one must not only believe A, but also believe that A is a reason for believing B, then in order for the belief that A together with the belief that A is a reason for B to be a reason for believing B, one would also have to believe that these two beliefs constitute a reason for believing B. An infinite regress results, of the very sort alluded to earlier. Although forming beliefs about reasons is one way in which one might prove to be sensitive to reasons, it is not the only way. As the infinite regress argument shows, it could not be the only way. When my dog forms the belief that there is food in his bowl as a result of hearing the food making a characteristic sound as it strikes the metal of the bowl, the belief he forms is responsive to reason: the

[26] See, for example, (Wilson 2002). For a different approach, but one which makes second-order cognition no less metaphysically tractable, see (Ericsson and Simon 1993).

characteristic sound provides him with good reason to believe that there is food in his bowl, and he is demonstrably sensitive to this very reason. The many animals that engage in problem solving show a remarkably complex sensitivity to reasons. They recognize whether various problem-solving strategies have succeeded or failed, and, when they have failed, they are responsive to this failure as a reason to try another strategy. This kind of responsiveness to reason does not require conceptualizing reasons as reasons, even if it is true that such a complex conceptual ability may provide one with greater and more subtle sensitivity to reason.

One might, of course, suggest some sort of debunking interpretation of animals and young children. While my dog responds to the sound of food being poured into his bowl, there are many reasons he is unresponsive to. Even the reasons he does respond to are ones which he responds to imperfectly. For some, this will suggest that talk of reasons here is inapposite: these considerations show that the dog is not ever responding to reasons. Instead, when we think about the proper explanation of animal behaviour, we should simply adopt what Dennett calls 'the design stance'[27]: the dog was built, as it were, to respond in certain sorts of ways; he is simply responding in the way he was built to respond. No talk of reasons is really called for.

But if this is one's reason for adopting the debunking explanation in the case of animals and young children, then one will be forced to offer a debunking account of reason itself. Adult human beings are not sensitive to all the reasons there are for belief, and even the reasons we are sensitive to are ones we are sensitive to only imperfectly. Gamblers, notoriously, are terribly insensitive to all manner of reasons. And what is true of gamblers is true of the rest of us. When it comes to the debunking account of reasons, what's bad for the goose is bad for the gambler. Surely this argument proves too much.

One further point is worth making here. The suggestion that normative assessment of any kind, such as the assessment of beliefs as justified or unjustified, presupposes some sort of voluntary control is a suggestion which quite a number of authors have made, including many whose motivations are quite different from those of the authors discussed here.[28] The authors under discussion wish to tie epistemic assessment to

[27] See (Dennett 1978) especially Chapter 1. Dennett himself, of course, would reject this move.

[28] Thus, for example, see the discussion of this issue in (Alston 1988) and (Plantinga 1993: Ch. 2).

agency, and then argue that only human adults have the requisite kind of agency. But any such move is far too quick. As many authors have now pointed out, there are all manner of cases of normative assessment which presuppose nothing whatever about agency.[29] The special case of talk of reasons at issue here is no different.

If information processing in children and non-human animals is thought of as nothing more than an assemblage of the kinds of mechanisms we see in the fly-snapping of frogs and the nest preparation of *Sphex*, then there is a good case to be made for the suggestion that such creatures should not be thought of in cognitive terms, and thus should not be regarded as responsive to reason. But as we have seen, cognition in animals and children is not at all like this, and the motivation for a debunking account of their apparent responsiveness to reason is undermined. Responsiveness to reason comes in degrees. Just as individual adult humans show a range of responsiveness to reason, different kinds of creatures show a range of such responsiveness. Animals who learn thereby demonstrate reasons responsiveness, and those who demonstrate sophisticated problem-solving skills manifest quite subtle and skilful patterns of responsiveness to reasons. Conceptualization of reasons qua reasons, and the ability to reflect on one's own cognitive states and processes, manifests a greater cognitive sophistication still, but it is no prerequisite for sensitivity to reason, and thus no prerequisite for the aptness of normative assessment.

Claim 7—that only adult human beings are capable of genuine knowledge—is thus undermined as well. If we reject the suggestion that only adult human beings are fit subjects for normative assessment, as I have argued we must, then the basis for restricting talk of knowledge to such adults is thereby undermined, unless we endorse the still stronger and far more controversial claim—Claim 8—that only adult human beings have beliefs.

So what is to be said on behalf of the claim that we should restrict the realm of believers to adult human beings? As John McDowell, who endorses this view, acknowledges, this does present a problem in explaining how it is that every one of us succeeded in making the cognitive transition from childhood to maturity. As McDowell comments,

[29] For discussions of epistemic assessments which make this point, while disagreeing about much else, see (Feldman 1988, 2000, 2001) and (Kornblith 2001).

> Now it is not even clearly intelligible to suppose a creature might be born at home in the space of reasons. Human beings are not: they are born mere animals, and they are transformed into thinkers and intentional agents in the course of coming to maturity. This transformation risks looking mysterious. (McDowell 1996: 125)

It is worth expanding on this point. As I noted above, the standard account offered by developmental psychologists has it that children have beliefs, but, prior to approximately age four, they have no beliefs about their own mental states, and certainly no beliefs about reasons qua reasons, or about truth. The concept of an enduring object is one which they have quite early on, and they form a robust set of beliefs about what is going on in their environment. It is only after coming to understand a great deal about the physical world around them that they start to have the conceptual sophistication required to form beliefs about their own mental states. Prior to achieving this more advanced conceptual sophistication, however, a great deal of learning goes on. (Remember: we are talking about children prior to the age of four. Late in this period, they have not only learned a great deal about the world around them; they are also talking in extremely sophisticated sentences.) The kind of sophisticated learning and problem solving which goes on during this period is not only most naturally described in terms of the acquisition of beliefs, often by way of fairly complicated inference; no one has ever offered an account of such learning in any other terms.

If we insist, however, that we will not call anyone a believer until they have the more sophisticated conceptual scheme which begins to emerge at age four, then we will need to redescribe the learning that goes on prior to that point without talking about belief acquisition or inference. Now some authors, at this point, start talking about 'proto-beliefs' or other sorts of primitive mental representations,[30] as if this simple terminological manoeuvre allows us to avoid the problem. But the strategy of renaming these states is not a solution to the problem, for the fact remains that there is no motivation in the phenomena themselves for regarding the kind of information processing which goes on prior to age four as different in kind than the kind that goes on afterward. More concepts are added to the child's repertoire, but the kinds of mechanisms and states by way of which information is registered and processed remain the same.[31]

[30] See, for example, (Brandom 1998: 391).

[31] I have developed this argument in greater detail in (Kornblith 2007).

So what is McDowell's solution to this problem? If we insist that young children do not have beliefs and are not sensitive to reasons, how exactly do we account for the ever-increasing cognitive sophistication that we see during this period of pre-belief? And how do children make the transition from being creatures utterly lacking in beliefs and insensitive to reason to being creatures who do have beliefs and are sensitive to reason? Here is what McDowell says:

> A mere animal, moved only by the sorts of things that move mere animals and exploiting the sorts of contrivances that are open to mere animals, could not single-handedly emancipate itself into possession of understanding. Human beings mature into being at home in the space of reasons or, what comes to the same thing, living their lives in the world; we can make sense of that by noting that the language into which a human being is first initiated stands over against her as a prior embodiment of mindedness, of the possibility of an orientation to the world. (McDowell 1996: 125)

So it is language learning, according to McDowell, that allows for this transition. Now there are a number of problems with this. First, as I just noted, children learn a language prior to having the very concepts which McDowell and others see as a prerequisite for having beliefs. So the suggestion that language learning somehow ushers the child into the space of reasons seems to give the child beliefs just a bit too soon. Second, and more importantly, this suggestion seems to ignore, rather than address, the very real problem of accounting for all of the learning which goes on prior to the alleged acquisition of beliefs. And finally, the suggestion seems to be a failure even on its own terms. Even if we suppose, with McDowell, that there are no beliefs prior to language acquisition, and that the child is initiated into the space of reasons in the very acquisition of language, we have merely relocated the problem rather than solved it: what was once a problem for the individual child now becomes a problem for the origin of language itself. If children are only able to acquire intentional states because they are raised in an environment in which they can be initiated into language, exactly how did the first languages come about, since, as McDowell would have it, there were no beliefs prior to the existence of language, but one cannot make the transition from being a creature without beliefs to being a creature with them without being embedded in a culture where language is spoken? If children can acquire beliefs only because they are enculturated by way of a 'prior embodiment of mindedness', then we seem to be committed to the existence of that prior embodiment all the way back in time. This is a rather high price to pay

for insisting that children do not have beliefs until they are conceptually quite sophisticated.

Any suggestion that we are forced somehow to make sense of this by the fact that animal information processing is nothing but the operation of causal mechanisms, and *therefore* cannot involve either reason responsiveness or belief,[32] surely runs afoul of the problem that it proves too much. Adult human belief acquisition is causally mediated; the science of psychology explains it, just as much as the information processing in children and non-human animals, by bringing it under the scope of psychological law. If this is sufficient to undermine the attribution of belief and sensitivity to reasons, then no one has beliefs at all. This may be congenial to eliminativists, but it was not where this argument was supposed to lead.

Finally, let me briefly address Claim 9, the suggestion that human cognition is not a natural phenomenon, that it may not be fully understood by being brought under the purview of the sciences, unlike the sort of information processing which goes on in other animals and young children. I believe the various claims which are meant to support this have already been addressed: that the divide between adult humans and others tracks the distinction between creatures who can learn and those who cannot; that it also tracks the distinction between epistemic agents and those who are not; that it also tracks the distinction between creatures who are responsive to reasons and those who are not, as well as the distinction between creatures whose states are subject to normative assessment and those who are not; and, finally, that it also tracks the distinction between those who have knowledge and belief and those who do not. With all of these supporting claims undermined, we are left with no good reason at all to believe that human cognition is not a natural phenomenon, capable of being fully explained by the cognitive sciences. Given the great success of the enterprise of cognitive science, this should come as no surprise at all.

IV

I have tried to articulate the common core of a certain Kantian approach to human cognition, an approach which would remove human reason from

[32] Precisely this move is strongly suggested by the insistence that there is an important contrast to be drawn between the 'space of causes' and the 'space of reasons'.

the natural world and make it somehow immune to the progress of scientific understanding. I have argued that this approach makes a very large number of empirical presuppositions about the differences between the kinds of states and processes which are involved in non-human animal information processing as well as information processing in very young children, on the one hand, and the kinds of states and processes which are involved in adult human cognition, on the other. I have argued that every last one of the crucial empirical presuppositions of this view is false, and that there is thus very good reason to reject it. Human reason is well within the scope of the naturalistic project.[33]

References

Alston, W. 1988. The Deontological Conception of Epistemic Justification. *Philosophical Perspectives* 2: 257–99.

Astington, J., Harris, P., and Olson, D. Eds. 1988. *Developing Theories of Mind*. Cambridge: Cambridge University Press.

Brandom, R. 1994. *Making It Explicit: Reasoning, Representing and Discursive Commitment*. Cambridge, MA: Harvard University Press.

—— 1998. Insights and Blindspots of Reliabilism. *Monist* 81: 371–92.

—— 2000. *Articulating Reasons: An Introduction to Inferentialism*. Cambridge, MA: Harvard University Press.

Carroll, L. 1895. What the Tortoise Said to Achilles. *Mind* 4: 278–80.

Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.

—— 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Davidson, D. 1984. Thought and Talk. In his *Truth and Interpretation*: 155–70. Oxford: Oxford University Press.

—— 2001. Rational Animals. In his *Subjective, Intersubjective, Objective*: 95–105. Oxford: Oxford University Press.

Dennett, D. 1978. *Brainstorms: Philosophical Essays on Mind and Psychology*. Cambridge, MA: Bradford Books.

—— 1984. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, MA: MIT Press.

Ericsson, K. A. and Simon, H. 1993. *Protocol Analysis: Verbal Reports as Data*, revised edition. Cambridge, MA: MIT Press.

[33] I have received helpful comments on this paper from audiences at University College, London, Brandeis University, and referees with Oxford University Press.

Evans, H. E. 1996. *The Comparative Ethology of the Sand Wasps*. Cambridge, MA: Harvard University Press.

Feldman, R. 1988. Epistemic Obligations. *Philosophical Perspectives* 2: 235–56.

—— 2000. The Ethics of Belief. *Philosophy and Phenomenological Research* 60: 667–95.

—— 2001. Voluntary Belief and Epistemic Evaluation. In M. Steup (ed.), *Knowledge, Truth and Duty: Essays on Epistemic Justification, Responsibility and Virtue*: 77–92. Oxford: Oxford University Press.

Fodor, J. 1983. *The Modularity of Mind*. Cambridge, MA: MIT Press.

Gopnik, A. and Meltzoff, A. 1997. *Words, Thoughts, and Theories*. Cambridge, MA: MIT Press.

Haugeland, J. 1998. *Having Thought: Essays in the Metaphysics of Mind*. Cambridge, MA: Harvard University Press.

Kelley, D. 2004. Vocal Communication in Frogs. *Current Opinion in Neurobiology* 14: 751–7.

Kornblith, H. 1993. *Inductive Inference and its Natural Ground*. Cambridge, MA: MIT Press.

—— 2001. Epistemic Obligation and the Possibility of Internalism. In A. Fairweather and L. Zagzebski (eds.), *Virtue Epistemology: Essays on Epistemic Virtue and Responsibility*: 231–48. Oxford: Oxford University Press.

—— 2002. *Knowledge and its Place in Nature*. Oxford: Oxford University Press.

—— 2007. The Metaphysical Status of Knowledge. *Philosophical Issues* 17: 145–64.

—— 2010. What Reflective Endorsement Cannot Do. *Philosophy and Phenomenological Research* 80: 1–19.

—— Manuscript. The Myth of Epistemic Agency.

Korsgaard, C. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. 1965. What the Frog's Eye Tells the Frog's Brain. In W. McCulloch (ed.), *Embodiments of Mind*: 230–255. Cambridge, MA: MIT Press.

Marr, D. 1982. *Vision*. San Francisco: W. H. Freeman.

McDowell, J. 1996. *Mind and World, with a new introduction by the author*. Cambridge, MA: Harvard University Press.

Moran, R. 2001. *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

Plantinga, A. 1993. *Warrant: The Current Debate*. Oxford: Oxford University Press.

Povinelli, D. 2000. *Folk Physics for Apes*. Oxford: Oxford University Press.

Povinelli, D. and Eddy, T. J. 1996. What Young Chimpanzees Know about Seeing. *Monographs of the Society for Research in Child Development* 6: 1–152.

Reader, S. and Laland, K. Eds. 2003. *Animal Innovation*. Oxford: Oxford University Press.

Ristau, C. 1991. Aspects of the Cognitive Ethology of an Injury-Feigning Bird, the Piping Plover. In C. Ristau (ed.), *Cognitive Ethology: The Minds of Other Animals*: 91–126. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shettleworth, S. 1998. *Cognition, Evolution and Behavior*. Oxford: Oxford University Press.

Skinner, B. F. and Ferster, C. 1957. *Schedules of Reinforcement*. New York: Appleton Century Crofts.

Tomasello, M. 2003. Chimpanzees Understand Psychological States: The Question is which Ones and to what Extent? *Trends in Cognitive Science* 7: 153–6.

Tomasello, M. and Call, J. 1997. *Primate Cognition*. Oxford: Oxford University Press.

Williams, M. 2004. Is Knowledge a Natural Phenomenon? In R. Schantz (ed.), *The Externalist Challenge*: 193–209. Berlin: de Gruyter.

Wilson, T. 2002. *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, MA: Harvard University Press.

Wooldridge, D. 1963. *The Machinery of the Brain*. New York: McGraw-Hill.

7

Naturalism, Transcendentalism, and Therapy

Penelope Maddy

The type of naturalism I'd like to explore here is a broadly methodological variety I call 'Second Philosophy'.[1] The leading theme of methodological naturalism, at least of the sort I'm after, is roughly that science is the best way we have of finding out about the world. If we could define 'science', this rough version would be enough—trust only the methods of science, or something like that—but we can't definitively circumscribe the methods of science, and it seems contrary to the ever-evolving, ever-improving character of that practice even to attempt this. In place of direct definition, then, my plan is to introduce an idealized enquirer called the Second Philosopher and to illuminate her character by comparing and contrasting her ways of proceeding with those of more familiar thinkers.

As my title suggests, the central foils will be transcendentalism—represented primarily by Kant—and therapeutic philosophies—including but not starting or ending with Wittgenstein. My favoured diagnostic instrument will be radical scepticism about the external world; the compare-and-contrast will focus on reactions to this familiar epistemological worry. I begin by sketching the character of the Second Philosopher and the relevant sense of First Philosophy. With this background in place, we take up the compare-and-contrast with Kant and the therapeutic philosopher in turn. I promise no startling theses by the end, but do hope that this

[1] The intended contrast is with ontological forms of naturalism. For more on Second Philosophy, see (Maddy 2007).

round-about discussion will at least bring the outlines of Second Philosophy into sharper focus.

I. The character of the Second Philosopher

The Second Philosopher is actually a quite mundane and familiar figure. She begins her investigations of the world with perception and common sense, gradually refines her observations, devises experiments, formulates and tests theories, always striving to improve her beliefs and her methods as she goes along; at some points in her investigation of the world, she addresses (her versions of) traditional philosophical questions; and the result is Second Philosophy. Unlike some near relatives, the Second Philosopher is simply born native to her particular point of view; she doesn't begin somewhere else, with certain apparently extra-scientific questions, and then turn to science for one reason or another,[2] a route that would seem to require a developed sense of what counts as 'science' and why it's to be preferred. Also, though the Second Philosopher obviously recognizes that there are people who employ methods different from hers that purport to get at what the world is like, she feels no temptation towards relativism: she straightforwardly explains why those alternative methods are ineffective.[3] In short, she believes as she does on grounds of the evidence, not because 'science' tells her so.

Employing our diagnostic, let's now ask how our Second Philosopher reacts to radical scepticism, in particular, to Barry Stroud's version of scepticism about the external world.[4] In aerial overview, this sceptic argues that I can't know I have hands unless I can know I'm not dreaming, and that I can't know I'm not dreaming because any test I might perform in an attempt to settle the question could itself be dreamt. The Second Philosopher might well hesitate over the subtleties of the notion of 'knowledge',[5] but setting these aside, if the sceptic has shown that it's no more reasonable

[2] Cf. (Quine 1975: 72): 'One [source of naturalism] is despair of being able to define theoretical terms generally in terms of phenomena.' See also (Fine 1996: 174): his goal 'is to situate humanistic concerns about the sciences within the context of ongoing scientific concerns, to reach out with our questions and interests to the scientist's questions and interests—and to pursue inquiry as a common endeavor'.

[3] Putnam thinks the naturalist is either a relativist or imperialist. In this terminology, the Second Philosopher is an imperialist. See (Maddy 2007: §I.7) for discussion and references.

[4] See (Stroud 1984, 2000). For more on what follows, see (Maddy 2007: §I.2).

[5] For example, does it require certainty of some kind? Does it have a performative, non-cognitive element? Is it determinate enough for these questions to have answers? And so on.

than not for her to believe she has hands, this is a serious challenge to her understanding of her cognitive abilities. She surely agrees that dream beliefs are generally unreliable, so the question is whether or not she has good reason now to believe that she's not dreaming, whether or not any test she could now perform would necessarily be useless.

The Second Philosopher's responses to questions like these may seem so pedestrian as not to count as proper entries in a serious philosophical debate. She will point out that dreams are never as continuous and coherent as her current experience, that she's now able to carry out a deliberate and sustained train of thought which she cannot do while dreaming, that she's familiar with a great body of anecdotal and experimental information about the nature of dreaming that informs her current opinion that she is awake. She reminds us that, with effort and training, people can successfully run tests while dreaming to confirm that they are in fact dreaming, in other words, that it isn't true that dreaming life and waking life are indistinguishable.[6] In a properly philosophical context, observations like these are shrugged off with the suggestion that all of this—the Second Philosopher's conviction that she's not dreaming, her sense of the continuity of her current state with her memory of the past and her intentions for the future, the very contrast she's been drawing between (apparent) dreaming and (apparent) waking—all this might be part of one much larger dream. At this point, the possibility being entertained isn't that I might be dreaming in the ordinary sense, but that I might be dreaming in an extra-ordinary sense that's the functional equivalent of the Evil Demon or the Brain in the Vat.

Fanciful hypotheses like these the Second Philosopher recognizes as colourful ways of posing a different kind of challenge. They're expressly designed to undercut all the ordinary evidence the Second Philosopher has been citing, all the ordinary methods she has for finding out about what the world is like, which means that the challenge isn't just to show that it's more reasonable than not to believe she has hands, but to show this 'from scratch' so to speak, that is, to show it without using any of her

[6] Common tests include looking at one's watch, especially more than once (for whatever reason, clock faces tend to look odd and the time they show doesn't stay constant); trying to jump (for whatever reason, jumping in dreams is exaggerated, like low-gravity jumping); attempting to put one's hand through an apparently solid object (for whatever reason, this is usually possible in a dream).

tried-and-trusted methods for showing things. This she doesn't know how to do, though she doesn't rule out in principle that there may be a way she hasn't thought of. The question, then, is whether or not her inability to meet this challenge implies that her belief in her hands, a belief based on her ordinary methods, is not reasonable after all. To put it another way: does her inability to defend her methods 'from scratch' show she can't reasonably believe them to be reliable?

The answer to this question appears to hinge on whether or not the 'from scratch' challenge arises inevitably from within the Second Philosopher's own ways of finding out about the world. The alternative would be that it arises only when one poses a peculiarly philosophical question about our knowledge of the world, when we feel the pull of a 'human aspiration . . . to get outside [our] knowledge and [our] condition' and to explain from this external perspective how '*any* knowledge of an independent world' is gained on *any* occasion.[7] The Second Philosopher can sympathize with this aspiration without holding that all claim to reasonable belief depends on its being satisfied. But if careful application of her very own methods leads to the conclusion that unless she can meet the 'from scratch' challenge, it's no more reasonable than not to believe she has hands—this would be a serious problem she couldn't safely set aside.

In fact, it would be the very problem once faced by the naturalistic Hume: he sets out to be the Newton of the Science of Man,[8] to found an 'experimental philosophy' of human nature based on 'experience' and 'a cautious observation of human life', but by the time he reaches the end of the first book of his *Treatise*, he finds himself 'ready to reject all belief and reasoning, and . . . look upon no opinion even as more probable or likely than another'; he is 'reduce[d] almost to despair . . . resolve[d] to perish' on this 'barren rock' (Hume 1739: Introduction, 7, 10; 1.4.7, 8, 1). Commentators continue to ponder how Hume managed to carry on his investigations after this disaster, but the Second Philosopher is puzzled by something else: wouldn't the natural response to such a shipwreck be a re-examination of the methods that led him there? And wouldn't such a re-examination suggest that something had gone wrong in his analysis of perception, perhaps in his argument to the conclusion that all we ever really or directly perceive are percepts, not external objects? If the Second

[7] The quotations are from (Stroud 2000: 138, 132).

[8] See (Stroud 1977: 5). For more, with references, see (Maddy 2007: §I.3).

Philosopher's own methods in fact led her to a similar shipwreck, she would diligently re-examine them, to figure out where she herself had gone astray.

Moore's famous appeal to common sense presents a contrast of a different kind. Moore purports to establish the existence of the external world by noting that 'here's a hand' and 'here's another'.[9] This is obviously no answer to the 'from scratch' question, because it appeals to our ordinary knowledge of our hands; in that respect, it resembles the Second Philosopher's first effort to respond to the dream argument, by citing various pedestrian facts. But we now understand the point of the Second Philosopher's insistence on ordinary dreaming: she's suggesting that the route to the 'from scratch' challenge in fact runs through extra-ordinary dreaming, that it doesn't arise directly from her familiar methods of finding out about the world, that those methods only require her to rule out that she isn't dreaming in the ordinary sense, something it's not particularly difficult to do. The contrast here is that Moore, at least in his 'Proof of an external world', doesn't feel the need to think about the sceptic's actual argumentation at all. In other places, he suggests that it needn't concern him because it begins from premises less certain than his belief in his hands.[10] This may well be so; in any case, the Second Philosopher wouldn't dispute it. She takes interest in the dream argument not because she doubts she has hands, but because a cogent argument from her ordinary methods to the 'from scratch' challenge would mandate a critical re-examination of those ordinary methods. Moore is apparently unconcerned with this possibility.

Finally, what about Descartes himself?[11] Contrary to his popular image, Descartes continues to regard ordinary beliefs as 'highly probable opinions . . . still much more reasonable to believe than to deny' (Descartes 1641: 22), even after strong grounds for doubt have been introduced. He resorts to the Evil Demon hypothesis only as an aid to

> turn my will in completely the opposite direction . . . by pretending for a time that these former opinions are utterly false and imaginary . . . (Descartes 1641: 22)

[9] See (Moore 1939). At least this is one reasonable interpretation of what Moore is up to, see (Stroud 1984: Ch. 3). Baldwin (1990: Ch. 9), presents an opposing view.

[10] See (Moore 1919: 227–8, 1940: 226). Moore (1941) is an exception—there he takes on the dream argument directly—but Moore was dissatisfied with this piece, see (Moore 1959: 13), and even it ends with a comparative judgement that his starting point is 'at least as good' as the sceptic's (p. 251).

[11] See (Maddy 2007: §I.1) for more.

in other words, in order to implement his overarching Method of Doubt. The plan is that this Method will allow us to uncover absolutely certain truths on the basis of which we can then rebuild, in a more dependable way, the edifice of science. What's worth noting is that the Second Philosopher has no reason to object to this procedure: Descartes is out to improve his ways of finding out about the world; he has a new method that he proposes for doing this. The Second Philosopher has no reason to think her current methods are all the reliable methods there are, so she's willing to give Descartes' proposal a try. Her disagreement with him comes only in the execution of this programme; she thinks he doesn't give good evidence for some of the principles he employs there.

What's important for our purposes is that the Second Philosopher doesn't dismiss Descartes' project as 'unscientific', as some of her near relatives might. She understands him as engaging in a serious attempt to find out about the world, just as she is; the disagreement between them has the character of an intra-mural squabble about which methods are effective and which aren't, and ultimately about concrete matters of fact. So, if we're to identify a sharp contrast class of First Philosophers, odd as it may seem, Descartes doesn't qualify.[12] As an example of someone who does qualify, let's consider the Constructive Empiricism of Bas van Fraassen.[13]

Van Fraassen holds that no evidence whatsoever could ever confirm the existence of unobservable entities. This isn't to say that we should banish them from our theories; the idea is rather that those theories aren't attempts to describe the world but attempts at empirical adequacy; that is, we don't aim for our theories to be true, but only for what they say about observable things to be true. Of course this won't sound right to the Second Philosopher, who's out to describe the world in the small as well as the large, and who thinks that the standard evidence, beginning with experiments on Brownian motion in the early 1900s, has established the existence of atoms; if van Fraassen maintains otherwise, she's eager to hear why he thinks this evidence is less than conclusive. Van Fraassen's surprising reply is that the Second Philosopher's evidence is just fine, that

[12] Despite the nearly overwhelming historical and rhetorical factors in favour of counting Descartes as the paradigm of a First Philosopher (to which I succumbed at the end of §I.1 and elsewhere in (Maddy 2007)), it now seems to me better terminology (closer to 'the joints') to use the term to mark the stark methodological contrast described here (and in (Maddy 2007: 61–2, 76, 85, and 308)). I'm grateful to Stroud (2009a) for prompting me to rethink this.

[13] For more on this topic, with references, see (Maddy 2007: §IV.1).

for her purposes it *is* conclusive, but that there are other purposes, philosophical or epistemological purposes, for which no evidence will do. On van Fraassen's picture, our investigation of what the world is like takes place at two distinct levels: the ordinary scientific level where the Second Philosopher resides and atoms are rightly said to exist; and the epistemic level where we step outside science and recognize that no evidence whatsoever could establish that atoms *really* exist.

From the Second Philosopher's perspective, this is simply baffling. She hasn't been told what's wrong with her evidence, but she is being encouraged to recognize that no evidence of this sort could ever confirm what she thinks it confirms, that in some sense or other, all of what she considers to be evidence is simply irrelevant to real question of the existence of atoms.

Here the similarity to the radical sceptic's 'from scratch' challenge is clear—the Second Philosopher is being asked to justify her belief in atoms without using any of the methods she has for justifying such things—except that the sceptic's challenge is relevant to the Second Philosopher, potentially casting doubt on the reliability of her methods, while van Fraassen's challenge is utterly irrelevant; it leaves her methods and beliefs entirely untouched. His sole complaint is that any methods within her grasp are only effective 'for the purposes of science', not 'for purposes of epistemology'. She will naturally wonder what the 'purposes of epistemology' are, and what methods are appropriate for those purposes, but nothing van Fraassen has to say about this is likely to convince her that a legitimate enquiry is involved. I reserve the term 'First Philosophy' for two-level views of this type.[14]

II. Kant's transcendentalism

With this rough sketch of First and Second Philosophy in place, let's now turn to Kant and begin by deploying our familiar diagnostic, the response to the sceptic. Kant's explicit treatment of the topic appears under odd terminology in the Refutation of Idealism:[15]

[14] Notice that the ill-chosen terminology of (Maddy 2007) invites the concern that the Second Philosopher can't both understand the sceptic and not understand the First Philosopher. The solution, obviously, is that the 'from scratch' challenge isn't a piece of First Philosophy (in the sense adopted here); if the Second Philosopher's contention that it doesn't arise directly from her methods can't be sustained, it poses a serious problem for her.

[15] (Kant 1781/7: B274–9), as supplemented by the long footnote in the B preface (Bxxxix–Bxli).

> Idealism . . . is the theory that declares the existence of objects in space outside us to be either merely doubtful and **indemonstrable**, or else false and **impossible**. (Kant 1781/7: B274)

The second of these—'dogmatic idealism' as Kant calls it—is meant to characterize Berkeley:

> who declares space, together with all the things to which it is attached as an inseparable condition, to be something that is impossible in itself, and who therefore also declares things in space to be merely imaginary. (B274)

This position is not the subject of the Refutation, however, as

> the ground for this idealism . . . has been undercut by us in the Transcendental Aesthetic. (B274)

The Refutation is addressed instead to the first, to 'problematic idealism',

> which . . . professes . . . our incapacity . . . for proving an existence outside us . . . by means of immediate experience. (B275)

This is the familiar external world scepticism we've been considering; Kant associates it with Descartes himself.[16]

The argument of the Refutation is breathtakingly simple on its surface: Kant claims that 'even our inner experience . . . is possible only under the presupposition of outer experience' (B275); in other words, based on facts about the nature of our immediate experience, we're to conclude that we know an external world. We have here the prototype of what's often called a 'transcendental argument'; P. F. Strawson famously attempted to reconstruct just such a line of thought.[17] Barry Stroud, in a sustained effort to assess arguments of this form, continues to 'cast doubt' on their prospects, 'especially when they are severed from the idealism that their success seems to depend on in Kant'.[18] This conclusion dovetails with the findings of leading interpreters of Kant, such as Henry Allison and Sebastian Gardner, according to whom Kant's Refutation presupposes his Transcendental Idealism.[19]

[16] One might wonder how Kant could have dealt with Berkeley's dogmatic idealism in the Aesthetic—presumably by establishing the empirical reality of space?—without having dealt with Descartes' problematic idealism at the same time. I take up this point below.

[17] See (Strawson 1959, 1966).

[18] See Chapters 2 (1968), 6 (1977), 11 (1994), 13 (1999), and 14 (1999) of (Stroud 2000). The quotation comes from the Introduction (p. xii).

[19] See (Gardner 1999: 179–96), and (Allison 2004: 300–3). At the crucial turn in the Refutation, Kant has shown that we must represent a thing outside us and concludes that this

From this general perspective, then, if we're to trace Kant's reply to the external world sceptic to its source, we must look to his case for Transcendental Idealism. This returns us to the Transcendental Aesthetic, where Kant intends to establish that space and time are merely forms of our intuition, and thus transcendentally ideal and empirically real.[20, 21] This short section of the *Critique* has inspired a tremendous literature over the centuries, but for present purposes, I hope it will suffice to divide these many subtle and diverse readings into two rough schools of thought.

Interpreters in the first of these schools see the argument for Transcendental Idealism as beginning in the Transcendental Exposition, from a premise concerning our knowledge of geometry: for example, Paul Guyer sees Kant as presupposing the necessity of our geometric knowledge; Waldemar Rohloff argues that the relevant premise is actually the *a priority* of applied geometry.[22] On this general style of interpretation—however the role of geometric knowledge is parsed in detail—Kant regards mathematics and pure science as part of our best theorizing about the world, as the most reasonable place to begin enquiry. If this is where the argument for Transcendental Idealism starts, and if Transcendental Idealism is presupposed in the Refutation of Idealism, then the Refutation isn't addressed to what we've been calling the sceptic's 'from scratch' challenge. Instead Kant is illustrating, within the context of the Transcendental Idealism he

must be 'a **thing** outside me and not . . . the mere **representation** of a thing outside me' (Kant 1781/7: B275). Gardner (1999: 185–6) explains: 'this inference cannot go through without some further assumption. "X exists" can be inferred from "X is a necessity of representation" . . . only . . . on the basis that X is a kind of thing the existence of which is tied to (a function of) necessities of representation.'

[20] The X of the previous footnote is a thing outside me, and thus a thing whose connection to the necessities of representation is established in the Aesthetic.

[21] It might be argued that there is an entirely independent case for Transcendental Idealism in the Transcendental Analytic, based on the role of the categories rather than the forms of intuition, and that this case is all that's needed for the Refutation. Such an independent argument is hard to find (see (Gardner 1999: 118, 120–5, 190–3), and even its conclusion is contentious (see (Bristow 2002) and (Watkins 2002) for relevant discussions). Fortunately, this topic can be set aside here: even if there is such an independent defence of Transcendental Idealism in the Analytic, the sketch in the two previous footnotes indicates that the Refutation rests at least in part on the ideality of space in particular, as argued in the Aesthetic. On the same grounds, I also set aside the possibility of what Ameriks calls 'a short argument' for Transcendental Idealism, that is, an argument that 'passes over Kant's own "long" and complex argument to idealism and its appeal to the specific features of our pure intuitions' (Ameriks 1992: 330).

[22] See (Guyer 1987: Ch. 16) and (Rohloff Unpublished).

recommends, how the sceptic goes astray, namely by assuming he can rely on his inner experience without presupposing outer experience.

The second school of interpreters traces the origins of the argument for Transcendental Idealism to the Metaphysical Exposition, where Kant argues that we have an *a priori* intuition of space. For example, Lisa Shabel argues that Kant first establishes his theory of space as an *a priori* intuition (in the Metaphysical Exposition), then shows how our geometric knowledge is based on that *a priori* intuition (in the Transcendental Exposition), and finally uses our geometric knowledge to build his case for Transcendental Idealism (in the Conclusions from the Above Concepts).[23] Given the mathematical and scientific developments since Kant, Allison prefers to skip the argument from geometry altogether, taking the line of thought to run directly from the *a priori* intuition of space (in the Metaphysical Exposition) to Transcendental Idealism (in the Conclusions) without the detour through the nature of our geometric knowledge (in the Transcendental Exposition).[24] Either way, these interpreters take the premises of the case for Transcendental Idealism to be the premises of the Metaphysical Exposition.

So, if we are to understand the structure of Kant's anti-sceptical line of thought on this second reading of the argument for Transcendental Idealism in the Aesthetic, we need to enquire into the presuppositions of the Metaphysical Exposition. Kant describes his starting point this way:

> By means of outer sense . . . we represent to ourselves objects as outside us, and all as in space. (A22/B37)

His plan is to 'expound the concept of space', which is to uncover 'that which belongs to' the concept (A23/B38), and this procedure is supposed to reveal its 'original representation' as an *a priori* intuition (A25/B40). So, for example, he argues that our representation of space cannot be acquired from experience, because

> In order for certain sensations to be related to something outside me . . . the representation of space must already be their ground. Thus the representation of space cannot be obtained from the relations of outer appearance through experience. (A23/B38)

[23] See (Shabel 2004).
[24] See (Allison 2004: Ch. 5).

Given the way Kant has set up the available options, this can be seen as an argument against Leibniz, for one.[25] It's a complex question to determine whether it weighs against Berkeley's account of how we come to construct spatial notions from our ideas,[26] but we needn't pursue this here; our concern isn't with Berkeley's empiricism, but with his subjective idealism.

As we've seen, the Berkeleian view that concerns Kant in the Refutation is his so-called 'dogmatic idealism'; this is the position Kant claims to have dealt with in the Aesthetic. What Kant seems to have in mind here is perhaps his own version of Berkeley, a figure for whom bodies are 'congeries of . . . ideas' (Berkeley 1713: 249) and therefore 'merely imaginary' (B274).[27] So it's natural to ask whether Berkeley (so understood) would accept the premise to the Metaphysical Exposition, that 'we represent to ourselves objects as outside us', in space. Presumably he *would* agree to this, *would* agree that we single out various batches of ideas as constituting objects other than ourselves, and indeed, as objects existing in space. More to the point for our purposes, Stroud's sceptic would easily agree to the same premise, that we represent objects as outside us in space. Are we to understand the argument of the Aesthetic as aiming to convince this sceptic of Transcendental Idealism, and hence to disprove, as in the Refutation, his problematic idealism? Or, to come at the matter more directly, does the Aesthetic aim to convince the sceptic that bodies don't just 'seem to exist outside me' (B69), that they are not 'mere illusion' (ibid.), that space is empirically real?[28]

Here I think we need to revisit the premise all have agreed to and ask if all have agreed to it in the same sense. Berkeley can say that 'we represent objects as outside us in space', but at least as Kant understands him, the sense in which he means this amounts to mere seeming. Similarly, the sceptic only affirms an apparent externality. If the premise is understood in

[25] See, for example, (Allison 2004: 101–3).

[26] See, for example, (Hatfield 1990: Ch. 3), (Falkenstein 1995: 174–5), and the references cited there.

[27] This ignores Berkeley's own claim to 'speak with the vulgar' (Berkeley 1710: §51) and to 'vindicate common sense' (Berkeley 1713: 244), as when Philonous remarks, 'I am of a vulgar cast, simple enough to believe my senses, and leave things as I find them. To be plain, it is my opinion, that the real things are those very things I see and feel, and perceive by my senses' (Berkeley 1713: 229). Margaret Wilson (1971: e.g. 468) describes Berkeley as laying claim to empirical realism.

[28] Again (as in footnote 16), if the Aesthetic eliminates Berkeley's dogmatic idealism, shouldn't it also eliminate Descartes' problematic idealism?

a weak sense that Berkeley or the sceptic would accept, then presumably the *a priori* intuition established at the conclusion of the Metaphysical Exposition would be of a purportedly external space, not of actually external space.[29] For Kantians in our second school of thought, like Allison and Shabel, the output of the Metaphysical Exposition feeds into a later argument for Transcendental Idealism, so we should ask whether the weaker output is sufficient to drive the later argument.

In Allison's version, the later argument in the Conclusions aims 'to determine what [an *a priori* intuition] could contain or present to the mind' (Allison 2004: 123); the goal is to understand the nature of the thing intuited, that is, the nature of space itself. For Shabel, the later argument of the Transcendental Exposition assumes that 'geometry is the science of space' and aims to answer the question

> how does our representation of space [that is, the *a priori* intuition guaranteed by the Metaphysical Exposition] manage to afford us those cognitions that are the unique domain of the science of geometry? (Shabel 2004: 202–3)

Either way, Kant takes the output of the Metaphysical Exposition to be an *a priori* intuition of real externality, not the bogus externality of the dogmatic idealist or the apparent-but-possibly-deceptive externality of the problematic idealist. Thus the argument must begin from a stronger premise than either idealist would allow. That the argument begins by assuming that our representations involve a robust externality is unproblematic if Kant's targets are (as he suggests)[30] Newton and Leibniz, who would both agree to this, but for our purposes, it is significant.

Still, we're left with an interpretive puzzle: if the Metaphysical Exposition begins from a strong version of the premise that we represent objects outside us in space, a version that Kant's Berkeley would not accept, why does Kant claim to have dealt with Berkeley in the Aesthetic?[31] Gardner offers a persuasive answer (Gardner 1999: 187–8). In the Refutation, Kant doesn't claim to have refuted dogmatic scepticism, but to have 'undercut' the 'ground' for it (B274). What is this ground? It is the notion that space is,

[29] Nothing in the argumentation would appear to signal a major shift of the sort that would be required otherwise.

[30] See (Kant 1781/7: A23/B37–8). Also see (Allison 2004: Ch. 5).

[31] Finally, the topic postponed in footnotes 16 and 28.

that it must be, 'encountered in things in themselves', what's often called 'transcendental realism' about space.[32] With striking sympathy, Kant writes:

> If one regards space and time as properties that, as far as their possibility is concerned, must be encountered in things in themselves, and reflects on the absurdities in which one then becomes entangled . . . then one cannot well blame the good Berkeley if he demotes bodies to mere illusion. (B70–1)

Once the possibility that space could be both transcendentally ideal and empirically real has been introduced, the sole motivation for Berkeley's empirical idealism is removed.

But our question here concerns Kant and the sceptic. We've traced the anti-sceptical argument of the Refutation back to the argument for Transcendental Idealism in the Aesthetic, and there we've found that even wildly divergent interpretations agree on a very general point: Kant begins with some store of common sense (we represent objects in space) and/or natural science (geometry describes the world). In this he roughly resembles Hume, Moore, and the Second Philosopher—none of these begins by addressing the 'from scratch' sceptic. Unlike Hume, Kant doesn't think his naturalistic starting point leads him to a sceptical outcome. Unlike Moore, he feels the need to address the sceptical argumentation, though he apparently isn't worried about the reliability of his initial beliefs:

> Geometry . . . follows its secure course . . . without having to beg philosophy for any certification of the pure and lawful pedigree of its fundamental concept of space. (B120)

> In mathematics and physics skepticism does not occur. (Kant 1800: 84)

If he's so confident on these matters, why does he bother with the Refutation at all? One plausible answer would be: because he wants to show that the 'from scratch' challenge doesn't arise if enquiry is pursued along his lines (in particular, that his methods don't undermine themselves by requiring an inference from inner to outer). In all this, then, Kant's approach to scepticism runs structurally parallel to the Second Philosopher's.

But Kant is anything but a Second Philosopher, and it's worth rehearsing why not.[33] When he examines his common sense/scientific starting

[32] See (Allison 2004: Ch. 2).
[33] Cf. (Maddy 2007: §I.4).

point, what strikes Kant is that some of this ordinary knowledge of the world is *a priori*. This aspect becomes his focus:

> I call all cognition transcendental that is occupied not so much with objects but rather with our mode of cognition of objects insofar as this is to be possible *a priori*. A **system** of such concepts would be called **transcendental philosophy** . . . This investigation . . . is to supply the touchstone of the worth or worthlessness of all cognitions *a priori*. (Kant 1781/7: A11–12/B25–6)

Of course, Kant doesn't regard his *a priori* cognitions as worthless. His explanation of their worth leads to the core of his philosophy—the Copernican Revolution:

> Let us once try whether we do not get farther . . . by assuming that the objects must conform to our cognition, which would agree better with the requested possibility of an *a priori* cognition of them, which is to establish something about objects before they are given to us. This would be just like the first thoughts of Copernicus, who, when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest. (Bxvi)

—and from there to the crucial separation of the empirical from the transcendental:

> Our expositions . . . teach the **reality** . . . of space in regard to everything that can come before us externally as an object, but at the same time the **ideality** of space in regard to things when they are considered in themselves . . . We therefore assert the **empirical reality** of space . . . though to be sure its **transcendental ideality**. (A27–8/B44)

At the empirical level of enquiry, we investigate an objective world of external, spatio-temporal objects, even inferring the existence of things we can't perceive on grounds of their causal connection to things we do perceive. At the transcendental level, we explore the conditions of our *a priori* cognition to explain our *a priori* knowledge of the empirical world. Ordinary empirical psychology could at best tell us how we cognize objects, not how they necessarily are; this transcendental enquiry is to tell us how we must cognize objects, and thus how the empirical world necessarily is.

Here the Second Philosopher remains firmly lodged at the empirical level. If Kant tells her that space is in some sense ideal, she will want to hear more about this, but Kant replies that for her purposes, for empirical purposes, space is real, just as she thinks it is—much as van Fraassen assured

her that for her purposes, for scientific purposes, atoms are real, just as she thinks they are. And just as she once wondered what van Fraassen's epistemic purposes were and how they were to be achieved, she now wonders what Kant's transcendental purposes are and how they are to be achieved. This profile qualifies Kant as a true First Philosopher, in the sense used here. Indeed, Kant was perhaps the first First Philosopher,[34] the originator of the two-level idea and the inspiration, direct or indirect, for many of its latter-day incarnations.

One salient difference between Kant and van Fraassen is that Kant gives a more substantive answer to the Second Philosopher's question of motivation than van Fraassen did: transcendental enquiry is to provide an explanation for our *a priori* knowledge of the world. Unfortunately, the Second Philosopher isn't much moved by this: she doesn't think physical geometry is *a priori*, and if there are things we know independently of experience, her first thought will be to seek an explanation in the structure of our cognitive apparatus and how it came to be as it is. For her, any question of *a priority* is an ordinary question about how human beings know the world; transcendental enquiry remains unmotivated, not to mention that it's by no means clear what sort of enquiry it would be and what methods would be appropriate and reliable for its pursuit.[35]

III. Therapeutic philosophy

At this point, let's set the Aesthetic and the Analytic aside, and consider for a moment what goes on in the Transcendental Dialectic, where Kant examines our natural temptation towards transcendental illusion.[36] This isn't ordinary empirical illusion, as in optical illusions, but rather

> **Transcendental illusion**, which influences principles whose use is not ever meant for experience . . . but which instead, contrary to all the warnings of criticism, carries us away beyond the empirical use of the categories, and holds out to us the semblance of extending the **pure understanding** . . . to lay claim to a wholly new territory. (Kant 1781/7: A295–6/B352)

[34] I know of no earlier examples, but would be grateful to be corrected.

[35] To be fair to Kant, another important motivation for the transcendental level comes from his practical philosophy. The complete Second Philosopher would owe an account of morality and value.

[36] Here I follow Gardner (1999: Ch. 7), and Allison (2004: Chs. 11, 13, and 15).

What causes this special type of illusion?

> In our reason . . . there lie fundamental rules and maxims for its use, which look entirely like objective principles . . . the subjective necessity of a certain connection of our concepts . . . is taken for an objective necessity, the determination of things in themselves. (A297/B353)

This sort of illusion can't be removed entirely, any more than

> The astronomer can prevent the rising moon from appearing larger to him, even when he is not deceived by this illusion. (A297/B354)

Kant's goal, then, will be to 'uncover . . . the illusion . . . while at the same time protecting us from being deceived by it' (A297/B354), but he recognizes that

> even after we have exposed the mirage it will not cease to lead our reason on with false hopes, continually propelling it into momentary aberrations that always need to be removed. (A298/B354–5)

The patient won't be cured, but armed with the means to treat flare-ups of the chronic condition as they occur.

To see how this works, consider the idea of the spatio-temporal world.[37] As Kant sees it, the present moment is conditioned by the one before and the one before that, a region of space is conditioned by the space that bounds it and the space that bounds that; each of these generates a series of conditions for which our reason posits an absolute unconditioned: the spatio-temporal world as a whole (A411–13/B438–40). This unconditioned might be thought of as an endpoint to the series—the beginning of time, the limit of space—or as the whole series itself—the whole of time, the extent of space—which 'is always unconditioned, because outside it there are no more conditions regarding which it could be conditioned' (A417/B445). On the first picture, the world has a beginning in time and is bounded in space; on the second, the world has no temporal beginning and is spatially unbounded. In the First Antinomy, Kant presents metaphysical arguments that purport to show both options—thesis and antithesis—to be self-contradictory (A426–33/B454–61).

[37] This may sound like precisely what the Second Philosopher regards as the subject of her enquiries, but Kant's concern is with speculative metaphysics, not with empirical investigations like hers.

Presumably this is enough to show that we've encountered some kind of illusion, but what is the diagnosis and what is the treatment? The source of the trouble, according to Kant, is the very idea of the world as a spatio-temporal whole.[38] Empirically, we recognize the relevant progressions from conditioned to condition, but reason adds to this the idea of the completion of the series. This idea of the spatio-temporal world as a whole is

> either **too big** or **too small** for every **concept of the understanding** . . . For **if it is infinite** and unbounded, then it is **too big** for every possible empirical concept. **If it is finite** and bounded, then you can still rightfully ask: What determines this boundary? . . . Thus a **bounded** world is **too small** for your concept. (A486–7/B514–15)[39]

The fault lies with this idea of reason—the absolute unconditioned—which goes beyond experience:

> With all possible perceptions, you always remain caught up among **conditions**, whether in space or in time, and you never get to the unconditioned, so as to make out whether this unconditioned is to be posited in an absolute beginning of the synthesis or in the absolute totality of the series without a beginning . . . The absolute whole . . . the world-whole . . . has nothing to do with any possible experience. (A483/B511)

In attempting to make judgements about the world-whole, we are, as advertised, attempting to apply the understanding beyond its proper use.

The arguments of the antinomy depend on the implicit and seemingly unassailable assumption that the world is, spatially and temporally, either infinite or finite. But what if the very idea of the world is 'an empty and merely imagined concept' (A490/B518)? Then the argument dissolves:

> If one regards the two propositions, 'The world is infinite in magnitude', 'The world is finite in magnitude' as contradictory opposites, then one assumes that the world . . . is a

[38] Cf. (Kant 1781/7): A490/B518, 'Thus we have been brought at least to the well-grounded suspicion that the world-idea, and all the sophistical assertions about [it] that have come into conflict with one another, are perhaps grounded on an empty and merely imagined concept of the way the object of [this idea] is given to us; and this suspicion may already have put us on the right track for exposing the semblance that has long misled us.'

[39] See (Kant 1781/7: A486–7/B514–15) for the corresponding passage on time, '[If] **the world has no beginning**; then it is too **big** for your concept; for this concept, which consists in a successive regress, can never reach the whole eternity that has elapsed. [If] **it has a beginning**, then . . . it is **too small** for your concept of understanding in the necessary empirical regress. For since the beginning always presupposes a preceding time, it is still not unconditioned, and the law of the empirical use of the understanding obliges you to ask for a still higher temporal condition, and the world is obviously too small for this law.'

thing in itself . . . But if I take away this presupposition, or rather this transcendental illusion, and deny that it is a thing in itself, then the contradictory conflict between the two assertions is transformed . . . because the world does not exist at all . . . It is only in the empirical regress of the series of appearances . . . it is never wholly given, and the world is thus not an unconditioned whole, and thus does not exist as such a whole, either with infinite or finite magnitude. (A504–5/B532–3)

Here Transcendental Idealism is the key to resolving the antinomy.

Thus the diagnosis: reason is ever tempted to extend the conditioned to an absolute unconditioned, and to imagine it has thus constituted an object of cognition. If this is right, why the constant temptation? The final piece to the puzzle is that this tendency of reason has its beneficial aspect:

The transcendental ideas are never of constitutive use, so that the concepts of certain objects would thereby be given, and in case one so understands them, they are merely sophistical . . . however, they have an excellent and indispensably necessary regulative use, namely that of directing the understanding to a certain goal. (A644/B672)

This goal is the illusory one set by the unconditioned. Though we can never cognize the world-whole, we can keep it before us as a '*focus imaginarius*':

i.e., a point from which the concepts of the understanding do not really proceed, since it lies entirely outside the bounds of possible experience [but which] nonetheless still serves to obtain for these concepts the greatest unity alongside the greatest extension. (A644/B672)

The empty idea of the world-whole serves to regulate enquiry, to encourage us to ever-broader understanding of the empirical world in space and time. These very 'fundamental rules and maxims' (A297/B353, cited above) are what generate 'transcendental illusion' when they are erroneously taken for objective principles. Kant insists that this 'indispensably necessary' illusion 'can be prevented from deceiving' when we understand how an empty goal serves a valuable regulatory function (A644–5/B672–3). The illusion will never leave us—we will continue to pursue 'false hopes' and to suffer from 'momentary aberrations'—but Kant's treatment arms us when they 'need to be removed' (A298/B355, cited above).

I've used here the language of diagnosis and treatment because this aspect of Kant's philosophy can be seen as an early example of what's now

called 'therapeutic philosophy':[40] rather than arguing for a particular position in the controversy over the extent of space and time, Kant turns his attention to the participants in that controversy; he diagnoses them as subject to a certain kind of illusion, which he then traces to its sources; he suggests that a clear understanding of how the illusion arises, combined with proper vigilance, will liberate these philosophers from their empty and unproductive squabble. Kant intends to show that the very question they're out to answer is a bad one, however tempting it may be.

With Kant's example in mind, let's return to the topic of scepticism, and in particular, to a more recent therapeutic approach that appears in Carnap's 'Empiricism, Semantics and Ontology' (Carnap 1950). Here Carnap encourages the sceptic and his opponent to be 'tolerant in permitting linguistic forms' (Carnap 1950: 257). The idea is that questions of reality can only be posed and answered inside a linguistic framework, that a choice between linguistic frameworks isn't a matter of truth or falsity but of efficiency and fecundity. The debate over our knowledge of the external world concerns just such a choice: the sceptic advocates a linguistic framework with evidential rules too weak for the existence of his hands to be confirmed; his opponent advocates a framework with stronger evidential rules that do allow him to confirm the existence of his hands; both imagine that the question of which rules are correct has an objective answer. But, Carnap insists, there is no fact of the matter about which framework is correct—there are no framework-independent facts—the issue is just which framework is more effective for the purposes at hand. If the purposes are those of scientific enquiry, presumably the language with the stronger rules is pragmatically preferable, but this isn't to say that it's 'true' in any sense.[41] If the leading purpose is to avoid error at all costs, then

[40] Kant sometimes uses this language himself, e.g.: '*The Critique of Pure Reason* is a preservative against a malady which has its source in our rational nature. The malady is the opposite of the love of home (the home-sickness) which binds us to our fatherland. It is a longing to pass out beyond our immediate confines and relate ourselves to other worlds' (Reflexion 5073 (1776–8): Ak 18:79), as translated by Kemp Smith (1923: lv). I'm grateful to Jeremy Heis for calling this passage to my attention.

[41] See (Carnap 1950: 244), 'The thing language in the customary form [i.e. a language whose rules allow us to settle questions like "is there a white piece of paper on my desk?" in the affirmative by looking] works indeed with a high degree of efficiency for the purposes of everyday life. This is a matter of fact, based on the content of our experiences. However, it would be wrong to describe this situation by saying: "The fact of the efficiency of the thing language is confirming evidence for the reality of the thing world"; we should rather say instead: "The fact makes it advisable to accept the thing language".'

the sceptic's language is preferable. Once it becomes clear that the question at issue is ill-posed,[42] the conflict will dissolve at long last and the combatants will be free to move on to more productive pursuits.

Like the Second Philosopher, Carnap's philosophical therapist doesn't claim to have refuted the sceptic, to have located a false premise or an error of reasoning in the sceptic's bleak assessment of the prospects for meeting the 'from scratch' challenge. Instead, both find ways to set the sceptical question aside without answering it. This parallel is what I hope to explore; I wonder if Second Philosophy and therapeutic philosophy can coexist in mutually beneficial ways. Unfortunately, in the particular case of Carnapian therapy, I'm afraid the answer is clearly no. As with Kant's therapy, the Carnapian variety only works if we first accept a body of controversial theses: that there are no facts, no truths, no evidential relations outside of linguistic frameworks; that the choice between such frameworks is purely pragmatic. The Second Philosopher and the sceptic might well object that what counts as good evidence for what isn't a matter of conventional choice in this way; they might agree that adopting Carnap's tolerant attitude is an attempt, an ultimately ineffective attempt, to take a range of entirely legitimate questions off the table.[43] As therapists, both Carnap and Kant prescribe a considerable course of bitter medicine before any benefits can be gained, medicine the Second Philosopher at least will find unpalatable. Our hope is for a brand of therapy consistent with Second Philosophy, for a pure therapy that works without relying on such objectionable doctrines.

[42] It is what Carnap calls an 'external question', in this case the question '. . . of the reality of the thing world itself . . . this question . . . cannot be solved because it is framed in a wrong way. To be real . . . means to be an element of the system; hence this concept cannot be meaningfully applied to the system itself' (Carnap 1950: 243).

[43] Cf. (Richardson 1994: 80), 'Tolerance has a hidden agenda: to remove objections to the strongest possible systems of mathematics', and (Ricketts 2007: 219–20), 'In advocating Tolerance, Carnap urges that the proponent of a strong logic for science need not answer the objections of constructively minded mathematicians. Carnap's philosophy of mathematics does not justify his own logical preference [i.e., classical mathematics]; it does, however, as Alan Richardson has noted, remove objections to it. There is a parallel in Carnap's final attitude to the realism-idealism debate [i.e., the debate between the external world sceptic and his opponent]. Carnap thinks there is no well-defined question here; what is at issue is the choice of the form for an observation language in the language of science. In the protocol language debate, Carnap comes to favor use of a realistic observation language, one that speaks of observationally detectible qualities and relations of material bodies. His understanding of the realism-idealism debate removes idealistic [i.e., sceptical] objections to this choice.'

Of course the whole idea of therapeutic philosophy is primarily associated with the writings of Wittgenstein, so this is a natural place to look. I'd like to side-step the current lively debate over 'therapeutic' readings of the *Tractatus*, and begin by reviewing the brand of therapy presented in the *Philosophical Investigations*. This will set the stage for *On Certainty*, where scepticism is addressed most directly.

In well-known passages from the *Investigations*, Wittgenstein writes

> Philosophy is a battle against the bewitchment of our intelligence by means of our language. (Wittgenstein 1953: §109)

> The problems arising through a misinterpretation of our forms of language have the character of *depth*. They are deep disquietudes; their roots are as deep in us as the forms of our language. (Wittgenstein 1953: §111)

> The confusions which occupy us arise when language is like an engine idling, not when it is doing work. (Wittgenstein 1953: §132)

The goal of philosophy, then, is to remove these confusions, bewitchments, and disquietudes:

> The real discovery is the one that makes me capable of stopping doing philosophy . . . the one that gives philosophy peace . . . There is not *a* philosophical method, though there are indeed methods, like different therapies. (Wittgenstein 1953: §133)

> The philosopher's treatment of a question is like the treatment of an illness. (Wittgenstein 1953: §255)

This isn't an empirical or theoretical enquiry of any kind:

> Our considerations could not be scientific ones . . . we may not advance any kind of theory. There must not be anything hypothetical in our considerations. We must do away with all *explanation*, and description alone must take its place. And this description gets its light, that is to say its purpose, from the philosophical problems . . . they are solved . . . by looking into the workings of our language, and that in such a way as to make us recognize those workings: *in despite of* an urge to misunderstand them. (Wittgenstein 1953: §109)

Here we have the promise of a pure form of therapy, one that doesn't depend on any problematic theorizing, but simply frees us from the various mental cramps brought on by misunderstandings of how our language actually works: 'the philosophical problems should *completely* disappear' (Wittgenstein 1953: §133).

Much of what Wittgenstein says outside these self-consciously methodological passages can then be seen, to a first approximation, as analogous to the sort of thing the psychoanalyst says to the patient during their therapeutic sessions: not as assertions or even questions of the usual sort, but as provocations designed to induce the desired state of disengagement from philosophical perplexity. This suggests that it would be inappropriate to subject these utterances to the usual techniques of philosophical analysis and critique; Wittgenstein is merely 'assembling reminders' (Wittgenstein 1953: §127) of 'what we have always known' (Wittgenstein 1953: §109). But, as Brian Rogers (unpublished a) has observed, this doesn't exempt Wittgenstein from all critical analysis: the practice of a given form of therapy often rests on substantial theoretical underpinnings (think of Freud!); one doesn't assert these theses in the course of treatment, but they inform the choice of what one does say. The second-philosophically troublesome theorizing in Kantian and Carnapian therapy is overt, but we must be alert to the possibility that even apparently pure therapy might harbour controversial theorizing within its motivating assumptions.

So here's the worry: does Wittgenstein's therapeutic practice rest on a debatable meta-philosophical theory of what constitutes a philosophical problem? If so, might this meta-theory artificially block investigation of legitimate, important matters? Consider, for example, the question: what is the nature of logical truth? The Second Philosopher has addressed this question, or at least, questions nearby—if it's either red or green, and it's not red, why *must* it be green?—and she produces an ordinary, contingent empirical theory about the structure of the world and the facts of human cognition in her attempt to answer it.[44] But, given Wittgenstein's interests in the *Tractatus* and the *Investigations*, the nature of logical truth might seem a paradigm of the type of question he hopes to cure us of asking. Is the Second Philosopher attempting to dig beneath the bedrock?[45] Does Wittgenstein covertly presuppose a theory according to which she is doing so, a theory that would inform his approach to diagnosing and treating this purported illness of hers?

[44] See (Maddy 2007: Part III).

[45] See (Wittgenstein 1953: §217), 'If I have exhausted the justifications I have reached bedrock, and my spade is turned. Then I am inclined to say, "this is simply what I do".'

I think there's at least one strand of Wittgensteinian thought that allows us to answer this question in the negative.[46] The nature of logic is in fact a recurring example in the meta-philosophical sections of the *Investigations* (Wittgenstein 1953: §§89–133), but it appears in a particular guise; he writes

> In what sense is logic something sublime? (Wittgenstein 1953: §89)

> Thought is surrounded by a halo.—Its essence, logic, presents an order, in fact the a priori order of the world. . . . It must . . . be of the purest crystal. (Wittgenstein 1953: §97)

Wittgenstein presents us with two enquirers:

> One person might say 'A proposition is the most ordinary thing in the world' and another: 'A proposition—that's something very queer'. (Wittgenstein 1953: §93)

This second attitude is 'in germ the subliming of our whole account of logic' (Wittgenstein 1953: §94); it leaves us 'unable simply to look and see how propositions really work' (Wittgenstein 1953: §93). Now I take it that the Second Philosopher, unencumbered by any preconception about what logic must be like, is able to do just this—look and see—while the encumbered enquirer 'is directed not toward phenomena, but . . . toward the "*possibilities*" of phenomena' (Wittgenstein 1953: §90), directed towards some elusive essence. For this enquirer, no empirical investigation is to the point:

> The more narrowly we examine actual language, the sharper becomes the conflict between it and our requirement. (For the crystalline purity of logic was, of course, not a *result of investigation*: it was a requirement.) (Wittgenstein 1953: §107)

This is the person Wittgenstein aims to cure:

> We have got onto slippery ice where there is no friction and so in a certain sense the conditions are ideal, but also, just because of that, we are unable to walk. We want to walk: so we need *friction*. Back to the rough ground! (Wittgenstein 1953: §107)

The Second Philosopher never left rough ground in the first place, so her investigation of logic is unproblematic; she stands in need of no cure.

[46] I'm grateful to Rogers for many conversations over the years on the general topic of Wittgenstein's attitude towards scientific investigations (of which the Second Philosopher's study of logical truth is one example). I think the line in the text matches the general outlines of what he's been urging on me for some time.

Thus we needn't see the therapeutic Wittgenstein as holding that certain questions are themselves wrong-headed, as long as they are asked in the right spirit. His goal is simply to treat those he finds, empirically, in a certain kind of predicament: he probes to discover if their inability to find the answers they seek may spring from their having set unexamined preconditions that keep the ordinary, empirical answers from satisfying them; if so, he then invites direct examination of those preconditions in the hope that they will dissolve in the patient's hands.

With this general understanding of the nature of Wittgensteinian therapy, let's turn to his extended discussion of scepticism in *On Certainty*. Many straightforwardly theoretical readings of the book have been given, mostly focused on the idea of indubitable hinge propositions.[47] Even therapeutic readings often tend towards overtly theoretical varieties of therapy.[48] The usual difficulties of interpreting Wittgenstein's writings are compounded in this case by the fact that these are first-draft notes; Wittgenstein died before he could rework and reassemble them into the sort of rich tapestry of interacting voices we find in the *Investigations*. Fortunately our interest here is localized: we want to explore the idea of a purely therapeutic response to the sceptic, so as to compare-and-contrast such therapy with Second Philosophy.

With this in mind, one immediately striking feature of *On Certainty* is that the patient most obviously up for treatment of this sort isn't the sceptic at all, but G. E. Moore![49] One prominent thread in this discussion starts early on:

> 'I know' often means: I have the proper grounds for my statement. So if the other person is acquainted with the language-game, he would admit that I know. (Wittgenstein 1969: §18)

The statement 'I know that here is a hand' may then be continued: 'for it's *my* hand that I'm looking at'. Then a reasonable man will not doubt that I know.—Nor will

[47] See (Moyal-Sharrock and Brenner 2005) for a selection.

[48] See, for example, (McGinn 2003), where Wittgensteinian therapy seems to involve getting us to understand the show/say distinction. In (Williams 2004), the therapy involves coming to see, among other things, that 'there are physical objects' is nonsense.

[49] Rogers (Unpublished b) suggests this is because there aren't actual sceptics around to treat, but there are people, like Moore, who attempt to refute the sceptic (though see footnote 9). The most conspicuous theme in the discussion of Moore arises from Malcolm's contention that Moore has misused the word 'know', see (Malcolm 1949). The evolution of this line of thought in Malcolm, Wittgenstein, and elsewhere, is fascinating, but beside the point here.

the [sceptic[50]]; rather he will say that he was not dealing with the practical doubt which is being dismissed, but there is a further doubt *behind* that one.—That this is an *illusion* has to be shewn in a different way. (Wittgenstein 1969: §19)

This Doubt behind the doubt can't be answered by straightforward appeal to what we ordinarily take ourselves to know; what's at issue here is the distinction between an ordinary question and the sceptic's 'from scratch' question.[51] If Moore is the patient, the therapist is pointing out to him that his removal of any 'practical doubt' isn't going to satisfy someone who's bothered by another sort of Doubt entirely:

If I don't know whether someone has two hands (say, whether they have been amputated or not) I shall believe his assurance that he has two hands, if he is trustworthy. And if he says he *knows* it, that can only signify to me that he has been able to make sure, and hence that his arms are e.g. not still concealed by coverings and bandages, etc. etc. My believing the trustworthy man stems from my admitting that it is possible for him to make sure. But someone who says that perhaps there are no physical objects makes no such admission. (Wittgenstein 1969: §23; see also §259)

There's no denying that Moore does sometimes seem strangely unresponsive to the sceptic's real concerns. The Wittgensteinian therapist's first step then is to get him to appreciate the nature of the question.[52]

But isn't there something odd about this? If Wittgenstein's goal is to free Moore from a philosophical perplexity, it's hard to see why he should begin by taking pains to induce that very perplexity. If Moore's temperament leaves him somehow immune to the sceptic's worries, if he feels content to offer ordinary answers to what's intended as an extra-ordinary question, then isn't he simply failing to feel the perplexity in the first

[50] Wittgenstein writes 'idealist' here, but this is apparently the figure we've been calling the 'sceptic'.

[51] De Pierris (1996: 181) argues that the distinction between 'philosophical and non-philosophical standpoints' is 'the most important theme of *OC* whereas scepticism is the vehicle that takes us through the journey'.

[52] In later sections of the book, written during a period when Wittgenstein met regularly with Moore, Wittgenstein seems convinced that Moore understands the sceptic's challenge and is pitching his response in the same register: 'When Moore says "I know that that's . . . " I want to reply "you don't *know* anything!"—and yet I would not say that to anyone who was speaking without philosophical intention. That is, I feel (rightly?) that these two mean to say something different' (Wittgenstein 1969: §407). 'When one hears Moore say "I *know* that's a tree", one suddenly understands those who think that that has by no means been settled' (Wittgenstein 1969: §481). 'It is as if "I know" did not tolerate a metaphysical emphasis' (Wittgenstein 1969: §482). Though it's well worth asking what Moore was really up to (cf. footnote 9), I won't address that question here.

place—and if so, why does he need any therapy at all? We might say the same of the Second Philosopher, who grasps the distinctive character of the 'from scratch' question, but doesn't see her inability to answer it as jeopardizing her ordinary beliefs. Using the case of logical truth as our guide, we might expect that the need for therapy arises when a thinker sets some precondition on what an answer to his question must look like, in this case, perhaps the requirement that a satisfactory answer to 'do you know?' must remove even the most hyperbolic doubt, must proceed, as we've described it, 'from scratch'. Given that neither Moore nor the Second Philosopher subscribes to this preconception, they would seem to be perfectly healthy to begin with, safely left to themselves.

One possible solution to this puzzle comes from Moore's understanding of his own philosophical project. In a well-known autobiographical passage, he writes:

> I do not think that the world or the sciences would ever have suggested to me any philosophical problems. What has suggested philosophical problems to me is things which other philosophers have said about the world or the sciences. (Moore 1942: 14)

This is borne out in his characterization of the problem he's addressing in 'Proof of an External World'. He begins by quoting Kant's 'scandal to philosophy' that the sceptic has not been answered, and announces in his own voice that

> There seems to me to be no doubt whatever that [this scandal] is a matter of some importance and also a matter which falls properly within the province of philosophy. (Moore 1939: 127)

Here Moore differs from the Second Philosopher, who draws no distinction between philosophical and scientific questions. She thinks it remains more reasonable than not to believe she has hands despite her inability to meet the 'from scratch' challenge, because that challenge doesn't arise in such a way as to undermine her ordinary methods; on these grounds, she explicitly rejects the relevant precondition, and thus needs no therapy. Moore, in contrast, isn't concerned with ordinary or scientific matters; he explicitly aims to address the philosopher's question. The trouble is that Moore, unlike the Second Philosopher, somehow fails to recognize or acknowledge the 'from scratch' character of the problem he undertakes to

solve. On this reading, then, the therapist first needs to get him to see the true nature of that question.[53]

Once Moore has been brought to share the sceptic's perplexity, we might then expect the Wittgensteinian therapist to isolate and dissolve the precondition that produces it. Perhaps this precondition is, as suggested, that a defence of our knowledge (or reasonable belief) must begin by answering the 'from scratch' challenge. The Second Philosopher has rejected this precondition on the grounds that the 'from scratch' challenge doesn't arise from straightforward application of her methods, but Wittgenstein's therapeutic philosopher would apparently attempt the more ambitious task of arranging for the patient to conclude that the precondition itself somehow dissolves on examination. (If the therapy is to remain pure, this must be accomplished without appeal to any controversial theorizing.) Here the Second Philosopher, lacking Moore's peculiarly 'philosophical' ambitions, will see little point in inducing false perplexity, and though she rejects the precondition in question, she doesn't see that there's anything incoherent about it, that it is in any sense an 'illusion'.[54]

[53] This may be the sort of thing Wittgenstein has in mind in his (1935: 108–9), 'As in the case of every philosophical problem, this puzzle arises from an obsession. Philosophy may start from common sense but it cannot remain common sense. As a matter of fact philosophy cannot start from common sense because the business of philosophy is to rid one of those puzzles which do not arise for common sense. No philosopher lacks common sense in ordinary life. So philosophers should not attempt to present the idealistic or solipsistic positions, for example, as though they were absurd—by pointing out to a person who puts forward these positions that he does not really wonder whether the beef is real or whether it is an idea in his mind, whether his wife is real or whether only he is real. Of course he does not, and it is not a proper objection. You must not try to avoid a philosophical problem by appealing to common sense; instead, present it as it arises with most power. You must allow yourself to be dragged into the mire, and get out of it. Philosophy can be said to consist of three activities: to see the commonsense answer, to get yourself so deeply into the problem that the commonsense answer is unbearable, and to get from that situation back to the commonsense answer. But the commonsense answer in itself is no solution; everyone knows it. One must not in philosophy attempt to short-circuit problems.' See also (Wittgenstein 1958: 58–9): 'There is no commonsense answer to a philosophical problem. One can defend common sense against the attacks of philosophers only by solving their puzzles, i.e. by curing them of the temptation to attack common sense; not by restating the views of common sense.' I'm grateful to Curtis Sommerlatte and Brian Rogers, respectively, for calling these passages to my attention.

[54] This may not be fair to Wittgenstein. I've characterized the 'from scratch' challenge in completely general terms—to justify belief in your hands without using any of your usual means of justification—and maintained that it's perfectly coherent. However, one might take the sceptic's challenge to be more specific than this—in particular, to require that the justification take the form of a cogent inference from knowledge of your inner states to

If this is the form of pure therapy Wittgenstein brings to the topic of scepticism, it appears to offer little of use to the Second Philosopher.

Still, the general idea that dissatisfaction with ordinary answers might spring from an unnoticed or unexamined precondition is a powerful one that the Second Philosopher might usefully deploy on suitable occasions—that is, when confronted with evidence that her fellow enquirers or even she herself is so encumbered. One example is the recent tendency towards a kind of ontological nihilism: given that the straightforward Quinean method of evaluating ontology has proved too simple, some recent philosophers conclude, on various grounds, that there is no objective way of settling ontological questions.[55] In contrast, the Second Philosopher believes her ordinary methods have established, for example, the existence of atoms. Why does this apparently respectable sort of ontological claim seem unacceptable to these philosophers? The answer, I've suggested, is that they're assuming, perhaps without having noticed, that the only proper way to judge ontology is by applying a certain kind of general criterion, a criterion in essential ways like Quine's, which they and the Second Philosopher agree is not likely to be forthcoming. Once this precondition is brought into the open, it simply lacks motivation and the Second Philosopher's piecemeal version of metaphysics naturalized appears entirely reasonable and viable.[56]

In any case, to return one last time to scepticism, I think we can find a more promising and straightforward brand of pure therapy in Austin's discussion of the Argument from Illusion in *Sense and Sensibilia*.[57] In the

knowledge of the external world—and there may be something incoherent about this. Cf. (Stroud 2009b).

[55] See (Maddy 2007: §IV.5), for more.

[56] Stroud (2009a) worries that the Second Philosopher would have no motivation to undertake the sort of therapy described here (and thus that I, the author of (Maddy 2007: § IV.5), am not a Second Philosopher). But mightn't a Second Philosopher want to encourage her fellow enquirers to address the important questions as she sees them? Wouldn't she be motivated to remove any obstacles to their joining with her in a cooperative effort?

[57] I should mention that in (Austin 1946) he confronts the sceptic more directly, with observations about the actual use of 'know' that appear to undercut his concerns. Though the degree of linguistic subtlety is considerably higher, this approach is akin to Malcolm and Wittgenstein's worries over Moore's use of the word. Grice later observed (see his (1989)) that inferences from facts of usage to the nature of knowledge have to be evaluated with some care, because a usage can be abnormal by violating conversational convention, without being false or meaninglessness (cf. Stroud's response to this part of Austin in his (1984: Ch. 2)). Wittgenstein clearly senses Grice's point in (Wittgenstein 1969: §§464, 552), and Grice

first five lectures,[58] Austin considers this familiar case for the philosopher's claim that we never directly perceive material objects, but only something else (sense data, percepts, ideas, impressions, . . .). He explicitly doesn't argue for a position of his own on perception;[59] he doesn't argue directly that one of the philosopher's premises is false or even meaningless in any theoretically loaded sense. Rather, his procedure is

> A matter of unpicking, one by one, a mass of seductive (mainly verbal) fallacies, of exposing a wide variety of concealed motives—an operation which leaves us, in a sense, just where we began. (Austin 1962: 4–5)

If there is a positive residue, it isn't a new philosophical thesis,

> But we may hope to learn something . . . in the way of a technique for dissolving philosophical worries (*some* kinds of philosophical worry, not the whole of philosophy). (Austin 1962: 5)

The central job is to unmask 'a certain special, happy style of blinkering philosophical English' (Austin 1962: 4).[60]

Anyone who's ever read this book will realize that I can't begin to summarize the wealth of observation and argumentation contained there, but I can give a quick listing to illustrate the kinds of warnings Austin sounds: looking for *the* kind of thing we perceive is already odd, given that we in fact perceive many different kinds of things (Austin 1962: 4, 7–8); the term 'material object' isn't an ordinary term but an undefined piece of jargon, introduced 'as a foil' for the equally undefined 'sense data' (Austin

(1989: 12–13) points out Austin's display of discomfort in the same vicinity (see also (Austin 1940: 64)). (This is part of the history set aside in footnote 49.)

[58] In lectures VI–IX, Austin discusses Ayer's linguistic understanding of sense-data; X–XI take up the question of incorrigibility.

[59] See (Austin 1962: 3–4), 'I am *not*, then—and this is a point to be clear about from the beginning—going to maintain . . . that we *do* perceive material things (objects). This doctrine would be no less scholastic and erroneous than its antithesis.'

[60] See (Fischer 2005) for a somewhat different take on (Austin 1962) as therapeutic. Because Ayer, the target of the lectures, doesn't truly believe the Argument from Illusion, Fischer sees 'exposing . . . concealed motives' as the leading method of the entire work. In contrast, it seems to me that the 'exposing' only begins when the issue of incorrigibility is raised (at the end of lecture IX), that the earlier discussion is aimed at undermining the appeal of the Argument from Illusion, and that 'unpicking . . . fallacies' is the method employed there. (I ignore the later parts because the philosopher I have in mind is Hume—see below—and whatever may have been the case with Ayer, I doubt Hume had any concealed motives.) In any case, Fischer's paper contains revealing analyses of some of Austin's 'unpicking', whether or not it's all done in the service of 'exposing'.

1962: 4, 7–8); not perceiving 'moderate-sized specimens of dry goods' (Austin 1962: 8) isn't the same as being deceived by one's senses (Austin 1962: 8–9); 'directly' in 'directly perceive' is used in some non-standard way that isn't specified (Austin 1962: 15–19); illusions are different from delusions, and conflating them facilitates the introduction of sense data (Austin 1962: 22–5); many commonly used examples are incompletely described or outright misdescribed (Austin 1962: 28–32); various 'delusive' experiences aren't in fact 'qualitatively indistinguishable' from ordinary experiences (Austin 1962: 48–50); things of quite different kinds can be 'qualitatively' similar (Austin 1962: 50–1), and so on. In general categories, Austin is counselling that we attend to how ordinary words are actually used, which is often considerably more subtle than we suppose; that we clearly define any new technical term, or new technical use of an ordinary term, and that we take care not to play illicitly on the ordinary meaning; that we guard against false dichotomies; that we look closely at the examples we use, to guard against over-simplifications, and so on. Here Austin isn't mounting a direct case against the Argument from Illusion; he's assembling an ever-growing tally of ignored complexities and problematic unclarities with the hope that we will gradually be released from the argument's grip, that its persuasiveness will slowly evaporate.

Now dissolving one argument commonly encountered on the road to the sceptical challenge obviously isn't enough to defeat that challenge completely, but given the particular roots of Hume's despair, for example, a course of Austinian therapy might well have done him a world of good! Perhaps the 'mainly verbal fallacies' that surround the theory of ideas and the Argument from Illusion are so 'seductive'—in both the thesis and the anti-thesis—that only concerted treatment is enough to root them out. And notice, also, that the illness here isn't a peculiarly philosophical strain, as it was for Wittgenstein and Moore; Hume is engaged in a straightforward empirical study of perception and human knowledge when he runs aground.

So now let's ask: is this therapy pure, or does it rely one way or another on theoretical presuppositions repugnant to Hume or the Second Philosopher? None of Austin's observations or general recommendations just canvassed appears to rest on overt theorizing of any kind, but perhaps they are informed by some covert theoretical assumptions. Here I think some would attribute to Austin the idea that the study of the ins and outs

of ordinary language is somehow all there is to philosophy, that the dictates of ordinary language are sacrosanct. Whatever might have been true of other so-called 'ordinary language philosophers',[61] this was not Austin's position. He clearly thinks that 'examining *what we should say when* . . . [is] at least . . . *one* philosophical method' and that 'evidently, there is gold in them thar hills' (Austin 1956a: 181), but he equally clearly insists that 'certainly ordinary language has no claim to be the last word', adding in his characteristic way 'if there is such a thing' (Austin 1956a: 185).

So what is Austin's method? It begins from a plain empirical claim:

> Our common stock of words embodies all the distinctions men have found worth drawing, and the connexions they have found worth marking, in the lifetimes of many generations: these surely are likely to be more numerous, more sound, since they have stood up to the long test of the survival of the fittest, and more subtle, at least in all ordinary and reasonably practical matters, than any that you or I are likely to think up in our arm-chairs of an afternoon—the most favored alternative method. (Austin 1956a: 182)

This last, obviously, is a dig at the philosopher's tendency to coin new terms, or to use old terms in new ways, without sufficient care. In *Sense and Sensibilia*, Austin writes:

> Tampering with words in what we take to be one little corner of the field is always *liable* to have unforeseen repercussions in the adjoining territory. Tampering, in fact, is not so easy as is often supposed . . . and is often thought to be necessary just because what we've got already has been misrepresented. And we must always be particularly wary of the philosophical habit of dismissing some of (if not all) the ordinary uses of a word as 'unimportant', a habit which makes distortion practically unavoidable. For instance, if we are going to talk about 'real', we must not dismiss as beneath contempt such humble but familiar expressions as 'not real cream'; this may save us from saying, for example, or seeming to say that what is not real cream must be a fleeting product of our cerebral processes. (Austin 1962: 63–4)

This isn't to say that technical terms shouldn't be introduced, only that this needs to be done with care. The concern is that, all too often, an ordinary term is being used in an extra-ordinary way (for example, 'direct', 'real') that remains unspecified because the word itself is a familiar one.

[61] As an example, Malcolm (1942: 349) argues that a philosophical statement is refuted by showing that it 'goes against ordinary language'. (Malcolm is attributing this position to Moore as well as endorsing it himself.)

Suppose, then, that we've managed to catalogue ordinary usage in our chosen domain and perhaps to have introduced some judicious and well-specified descriptive jargon; is this the end of the story? Not at all:

> Ordinary language . . . embodies . . . the inherited experience and acumen of many generations of men. But then, that acumen has been concentrated primarily upon the practical business of life. If a distinction works well for practical purposes in ordinary life (no mean feat, for even ordinary life is full of hard cases), then there is sure to be something in it, it will not mark nothing: yet this is likely enough not to be the best way of arranging things if our interests are more extensive or intellectual than the ordinary. (Austin 1956a: 185)[62]

In particular, Austin recognizes that

> [this inherited experience] has been derived only from the sources available to ordinary men throughout most of civilized history: it has not been fed from the resources of the microscope and its successors. (Austin 1956a: 185)

Presumably, if we wish to understand the physical constitution of matter, or the nature of biological processes, or the best way to build a bridge, we must go beyond the wisdom contained in ordinary language to the resources of scientific enquiry.[63] Indeed Austin notes that the study of ordinary language alone is most likely to be useful in cases where it is

> rich and subtle, as it is in the pressingly practical matter of Excuses, but certainly is not in the matter, say, of Time . . . In ordinary life we dismiss the puzzles that crop up about time, but we cannot do that indefinitely in physics. (Austin 1956a: 182, 186)

[62] See also (Austin 1956a: 185), 'ordinary language . . . in principle . . . can everywhere be supplemented and improved upon and superseded', and (Austin 1962: 63), 'Certainly, when we have discovered how a word is in fact used, that may not be the end of the matter; there is certainly no reason why, in general, things should be left exactly as we find them; we may wish to tidy the situation up a bit, revise the map here and there, draw the boundaries and distinctions rather differently.'

[63] In a biographical sketch, Warnock (1969: 4) notes Austin's background as 'a classical scholar and linguist', and continues: 'That this was his own training was, as he knew, significant for him; but he was very far from assuming, for that reason, that it was the best sort of training to have. It is possible that he himself would have preferred to be a scientist, and certain that he would have wished to know a great deal more about the sciences . . . his exacting habits of thought led him to question the value of the educational method by which he had acquired them, and he was sometimes inclined to think that he had wasted a great deal of time.' Cf. Urmson (1965: 83), quoting from Austin's notes: 'brought up on classics: no quarrel with maths etc., just ignorant'.

So one corrective to ordinary language is ordinary science. In his favoured case of Excuses, he lists several 'systematic aids': the dictionary,[64] the law, and psychology, anthropology, and animal behaviour (Austin 1956a: 186–9).[65]

If we follow Austin, then, if we apply the modes of enquiry he recommends and avoid the common pitfalls he warns against, what is the intended result? In the larger compass of Austin's work, the goal of these practices is straightforward—we're out to better understand the world:

> Words are not (except in their own little corner) facts or things: we need therefore to prise them off the world, to hold them apart from and against it, so that we can realize their inadequacies and arbitrariness, and can re-look at the world without blinkers . . . When we examine what we should say when, what words we should use in what situations, we are looking again not *merely* at the words (or 'meanings', whatever they may be) but also at the realities we use the words to talk about: we are using a sharpened awareness of words to sharpen our perception of, though not as the final arbiter of, the phenomena. (Austin 1956a: 182)

In the special case of the Argument from Illusion, this quest for better understanding takes a more therapeutic form: the cumulative effect of Austin's observations is aimed to get us over a certain mental tick.

Which brings us back to our question: does Austin's therapeutic treatment of the philosopher under the spell of the Argument from Illusion rest on any covert theorizing that the Second Philosopher would find problematic? It seems to me that the answer is no. The underlying belief that ordinary language often tracks real-world connections and distinctions is empirical; the approach to the study of ordinary language is empirical; the correctives imagined to the wisdom of ordinary language are empirical; even the test of the effectiveness of the therapy is presumably empirical.[66] If this is right, then we have here another example of pure therapy, albeit applied to only one portion of the sceptical problem. And as far as I can see, Austin's therapy is entirely congenial to the Second Philosopher; she

[64] Cf. (Urmson 1965: 79), 'Austin, who must have read through the *Little Oxford Dictionary* very many times, frequently insisted that this did not take so long as one would expect. '

[65] It's worth noting that Austin foresees efforts like his, in cooperation with those of grammarians and others, as eventually producing a true 'science of language' (Austin 1956b: 231–2).

[66] Cf. (Urmson 1969: 25), 'Austin regarded this method as empirical and scientific, one that could lead to definitely established results, but he admitted that "like most sciences, it is an art", and that a suitably fertile imagination was all important for success.'

might well have used it herself, faced with poor, despairing Hume (again a much more pressing case, from her perspective, than the peculiarly philosophical illness Wittgenstein diagnoses in Moore). Indeed, the whole of Austin's method seems to me to showcase a distinctive tool in the second-philosophical toolbox, one especially well-suited to the areas of enquiry he takes up.[67, 68]

Let me stop here. As predicted, I have no overarching conclusions to offer, but I hope to have clarified the character of Second Philosophy in a number of complementary ways, by sketching the Second Philosopher's response to radical scepticism and comparing it with those of Hume and Moore, and by specifying the relevant sense of First Philosophy. I've also suggested that Kant's 'refutation' of scepticism has more in common with the Second Philosopher's response than one might expect, despite his being perhaps the original First Philosopher, and I've indicated how and why the Second Philosopher remains unmoved by his motivations for transcendentalism. In the realm of therapeutic philosophies, we've seen how Wittgenstein aims to expose and dissolve the sort of unnoticed preconditions that can block our acceptance of ordinary answers to our questions; though the Second Philosopher may well doubt that all traditional philosophical problems take this form, she can welcome and practice this sort of therapy in appropriate cases. Finally, Austin's subtle approach to examining what-we-should-say-when promises a more direct, broadly applicable, down-to-earth form of therapy—and a new second-philosophical method for certain kinds of positive investigations as well. If, in all this, Second Philosophy has been clarified and perhaps some small

[67] For example, excuses, if and cans, performatives, etc. Cf. footnote 35.

[68] Wilson (2006) might be thought of as an application of Austin's method, with special attention to scientific terms like 'hardness'. Indeed Wilson writes: 'I will be flattered if the work is regarded as a worthy continuation of the school of tempered common sense pioneered by Thomas Reid and J. L. Austin' (Wilson 2006: xviii; see also 17–18). Ironically, when Wilson takes exception with Austin (Wilson 2006: 89, and especially 472–3), he seems to have missed what comes directly after the quotation above about 'the resources of the microscope', i.e., 'it must be added too, that superstition and error and fantasy of all kinds do become incorporated in ordinary language and even sometimes stand up to the survival test (only, when they do, why should we not detect it?)' (Austin 1956a: 185). Also, Wilson is a philosopher of language, intent on understanding how language correlates with the world; Austin, as we've seen, sees the study of language at least in part as a means of finding out about the world.

further interest sparked in its nature and practice, I confess that would be conclusion enough for me.[69]

References

Allison, H. 2004. *Kant's Transcendental Idealism*, revised and enlarged edition. New Haven: Yale University Press.

Ameriks, K. 1992. Kantian Idealism Today. *History of Philosophy Quarterly* 9: 329–42.

Austin, J. L. 1940. The Meaning of a Word. Reprinted in his 1979: 55–75.

—— 1946. Other Minds. Reprinted in his 1979: 76–116.

—— 1956a. A Plea for Excuses. Reprinted in his 1979: 175–204.

—— 1956b. Ifs and Cans. Reprinted his 1979: 205–32.

—— 1962. *Sense and Sensibilia*. Edited by G. J. Warnock. Oxford: Oxford University Press.

—— 1979. *Philosophical Papers*, third edition. Edited by J. O. Urmson and G. J. Warnock. Oxford: Oxford University Press.

Baldwin, T. 1990. *G. E. Moore*. London: Routledge.

Berkeley, G. 1710. *A Treatise Concerning the Principles of Human Knowledge*. Edited by J. Dancy. Oxford: Oxford University Press, 1998.

—— 1713. *Three Dialogues between Hylas and Philonous*. Edited by J. Dancy. Oxford: Oxford University Press, 1998.

Bristow, W. 2002. Are Kant's Categories Subjective? *Review of Metaphysics* 55: 551–81.

Carnap, R. 1950. Empiricism, Semantics and Ontology. Reprinted in P. Benacerraf and H. Putnam (eds.), *Philosophy of Mathematics: Selected Readings*: 241–57. Cambridge: Cambridge University Press, 1983.

De Pierris, G. 1996. Philosophical Scepticism in Wittgenstein's *On Certainty*. In R. Popkin (ed.), *Skepticism in the History of Philosophy*: 181–96. Dordrecht: Kluwer.

Descartes, R. 1641. Meditations on First Philosophy. In J. Cottingham et al. (eds.), *The Philosophical Writings of Descartes*, volume II: 3–62. Cambridge: Cambridge University Press, 1984.

Falkenstein, L. 1995. *Kant's Intuitionism*. Toronto: University of Toronto Press.

Fann, K. T. Ed. 1969. *Symposium on J. L. Austin*. London: Routledge.

Fine, A. 1996. Afterward to his *The Shaky Game*, 2nd edition: 173–201. Chicago: University of Chicago Press.

[69] Thanks to Jeremy Heis, Brian Rogers, Waldemar Rohloff, Lisa Shabel, James Weatherall, and the members of my 2007–8 seminar. It was a great pleasure to present and discuss this material at the May 2008 workshop on The Nature of Naturalism (part of an AHRC project on Transcendental Philosophy and Naturalism); my thanks to the organizers (especially Joel Smith) and the participants.

Fischer, E. 2005. Austin on Sense-Data: Ordinary Language Analysis as 'Therapy'. *Grazer Philosophische Studien* 70: 67–99.

Gardner, S. 1999. *Kant and the* Critique of Pure Reason. London: Routledge.

Grice, P. 1989. *Studies in the Way of Words*. Cambridge, MA: Harvard University Press.

Guyer, P. 1987. *Kant and the Claims of Knowledge*. Cambridge: Cambridge University Press.

Hatfield, G. 1990. *The Natural and the Normative*. Cambridge, MA: MIT Press.

Hume, D. 1739. *Treatise of Human Nature*. Edited by D. F. and M. J. Norton. Oxford: Oxford University Press, 2000.

Kant, I. 1781/7. *Critique of Pure Reason*. Translated and edited by P. Guyer and A. Wood. Cambridge: Cambridge University Press, 1997.

—— 1800. The Jäsche Logic. In J. M. Young (trans. and ed.), *Lectures on Logic*: 519–640. Cambridge: Cambridge University Press, 1992.

Kemp Smith, N. 1923. *A Commentary to Kant's* Critique of Pure Reason, 2nd edition. New York: Humanities Books, 1999.

Maddy, P. 2007. *Second Philosophy*. Oxford: Oxford University Press.

Malcolm, N. 1942. Moore and Ordinary Language. In P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*: 345–68. New York: Tudor.

—— 1949. Defending Common Sense. *Philosophical Review* 58: 201–20.

McGinn, M. 2003. Responding to the Skeptic: Therapeutic versus Theoretical Diagnosis. In S. Luper (ed.), *The Skeptics: Contemporary Essays*: 149–63. Aldershot: Ashgate.

Moore, G. E. 1919. Some Judgments of Perception. Reprinted in his *Philosophical Studies*: 220–52. London: Routledge, 1922.

—— 1939. Proof of an External World. Reprinted in his 1959: 127–50.

—— 1940. Four Forms of Skepticism. Reprinted in his 1959: 196–226.

—— 1941. Certainty. Reprinted in his 1959: 227–51.

—— 1942. An Autobiography. In P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*: 1–39. New York: Tudor.

—— 1959. *Philosophical Papers*. London: George Allen & Unwin.

Moyal-Sharrock, D. and Brenner, W. Eds. 2005. *Readings of Wittgenstein's* On Certainty. Basingstoke: Palgrave Macmillan.

Quine, W. V. O. 1975. Five Milestones of Empiricism. Reprinted in his *Theories and Things*: 67–72. Cambridge, MA: Harvard University Press, 1981.

Richardson, A. 1994. Carnap's Principle of Tolerance. *Proceedings of the Aristotelian Society, Supplementary Volume* 68: 67–82.

Ricketts, T. 2007. Tolerance and Logicism: Logical Syntax and the Philosophy of Mathematics. In M. Friedman and R. Creath (eds.), *Cambridge Companion to Carnap*: 200–25. Cambridge: Cambridge University Press.

Rogers, B. Unpublished a. Taking Wittgenstein Seriously as a Therapist. Unpublished work-in-progress.

—— Unpublished b. Wittgenstein's Philosophical Methods in *On Certainty*. Unpublished work-in-progress. (A shorter version appears in V. Munz, K. Puhl, and J. Wang (eds.), *Language and World*: 360–2. Kirchberg am Wechsel: Austrian Ludwig Wittgenstein Society, 2009.)

Rohloff, W. Unpublished. Kant's Argument from Geometry.

Shabel, L. 2004. Kant's 'Argument from Geometry'. *Journal of the History of Philosophy* 42: 195–215.

Strawson, P. F. 1959. *Individuals*. London: Routledge.

—— 1966. *The Bounds of Sense*. London: Routledge.

Stroud, B. 1977. *Hume*. London: Routledge.

—— 1984. *The Significance of Philosophical Skepticism*. Oxford: Oxford University Press.

—— 2000. *Understanding Human Knowledge*. Oxford: Oxford University Press.

—— 2009a. Review of *Second Philosophy*. *Mind* 118: 500–3.

—— 2009b. Scepticism and the Senses. *European Journal of Philosophy* 17: 559–70.

Urmson, J. O. 1965. A Symposium on Austin's Method. Reprinted in Fann 1969: 76–86.

—— 1969. Austin's Philosophy. In Fann 1969: 22–32.

Warnock, G. J. 1969. John Langshaw Austin, a Biographical Sketch. In Fann 1969: 3–21.

Watkins, E. 2002. Kant's Transcendental Idealism and the Categories. *History of Philosophy Quarterly* 19: 191–215.

Williams, M. 2004. Wittgenstein's Refutation of Idealism. In D. McManus (ed.), *Wittgenstein and Skepticism*: 76–96. London: Routledge.

Wilson, M. 2006. *Wandering Significance*. Oxford: Oxford University Press.

Wilson, M. D. 1971. Kant and 'the *Dogmatic* Idealism of Berkeley'. *Journal of the History of Philosophy* 9: 459–75.

Wittgenstein, L. 1935. *Wittgenstein's Lectures, 1932–1935*. Edited by A. Ambrose. Amherst, NY: Prometheus Books, 2001.

—— 1953. *Philosophical Investigations*, 3rd edition. Translated by G. E. M. Anscombe. Malden, MA: Blackwell, 2001.

—— 1958. *The Blue and Brown Books*. New York: Harper & Row, 1965.

—— 1969. *On Certainty*. Edited by G. E. M. Anscombe and G. H. von Wright, translated by D. Paul and G. E. M. Anscombe. New York: J. & J. Harper.

8

Is Logic Transcendental?

Peter Sullivan

1. Introduction

My intention here is to recommend a positive answer to the title question, on an understanding of it which makes this answer non-trivial, and which connects with recent discussion of problems facing arguments purporting to assign a transcendental status to their conclusions. The plan of approach is as follows.

I begin in §2 by presenting an 'easy' answer to the question. This answer is borrowed from a discussion of Thomas Nagel's, but one should not presume that when pulled from its context it can represent Nagel's own position.

In §§3–5 I offer an initial assessment of the easy answer, and consider why it can seem meagre and unsatisfying. First, in §3, the easy answer is assessed against a statement of Frege's about what logic is, and found not to deliver all we want. §4 then describes a setting—a 'post-metaphysical orientation'—in which the apparent weakness of the answer can begin to seem genuinely worrying. In §5 I consider and reject the suggestion that the strength of basic logical convictions could simply overpower those worries.

§§6–9 are concerned with a framework for transcendental argument developed by Mark Sacks, in his *Objectivity and Insight*, in which a more satisfying positive answer might be given. §6 first offers an overview of Sacks' broad strategy, by reference to two problems, the problem of 'the inference to reality' and the problem of 'universality', that are widely held to obstruct the project of transcendental argument. In §7 I adopt a somewhat

closer viewpoint, describing the three main planks of Sacks' construction, and raising some initial questions about them. §8 has the status of an interlude, and impressionistically sketches how Sacks' construction carries echoes not only of Kant, which one might expect, but also of Frege, which perhaps some wouldn't expect. Then in §9 two main questions are raised about how the construction meets the universality problem.

Finally §10 holds, first, that Sacks' framework does indeed yield a satisfying positive answer to our question; and second, that this conclusion calls for a reassessment of the 'easy' answer first considered.

It will emerge that a main difference between the unsatisfying and more satisfying answers to our question lies in how they understand the term 'transcendental'. The meaning of that part of the question therefore shifts in the course of the chapter. So, it would be as well to have an understanding of the other main term in the question that is solid and stable. Frege supplies this:

> the task we assign to logic is only of saying what holds with the utmost generality for all thinking, whatever its subject matter. We must assume that the rules for our thinking and for our holding something to be true are prescribed by the laws of truth. The former are given with the latter. Consequently we can also say: logic is the science of the most general laws of truth. (Frege 1979: 128)

2. The easy answer

According to one very common understanding the term 'transcendental' is regarded as merely a dependent abstract from the phrase 'transcendental argument'.[1] On that usage 'transcendental' applies to anything established by a transcendental argument.

On that same common understanding a transcendental argument is a variety of anti-sceptical argument which aims to refute scepticism from its own assumptions. The conclusion that such an argument intends to establish is something the sceptic 'chooses to doubt', for instance, that events are determined by their antecedents in accordance with causal laws. The aim is to show that this conclusion is presupposed by something else, the premise of the transcendental argument, that the sceptic would not, or perhaps even could not, dream of doubting, for instance, that we

[1] Bell laments this (1999: 193–4); I don't disagree, but am simply reporting the fact.

are aware of things happening. The argument will be neatly self-contained if the reason why the sceptic cannot doubt the premise of the argument is that this premise is involved in his formulation of his doubt about the intended conclusion, or even in the possibility of his formulating this doubt. This kind of self-containment is regarded as a particular merit of the transcendental strategy of argument.[2]

On this understanding it seems straightforward to argue that basic logical principles are transcendental. Scepticism about these principles must be more than a disregard or distaste for them. It must be an articulate, reasoned stance towards them. But to adopt any such stance is implicitly to commit oneself to the standards by which such stances are evaluated, and these minimally include the basic principles of logic. So, one cannot so much as formulate sceptical doubts about basic logical principles without refuting these doubts.

In *The Last Word*,[3] Thomas Nagel argues in just this way. He cites Descartes as someone who betrayed his own best insights by misguidedly trying to raise a variety of scepticism about logic. To do that, Nagel remarks, 'is itself an exercise of reason, and by engaging in it Descartes has already implicitly displayed his unshakeable commitment to first-order logical thought' (p. 60). He later makes the same point more entertainingly:

> suppose someone argues as follows (somewhat in the vein of Descartes' evil genius hypothesis):
>
> > If my brains are being scrambled, I can't rely on *any* of my thoughts, including the basic logical thoughts whose invalidity is so inconceivable to me that they seem to rule out anything, including scrambled brains, which would imply their invalidity—for the reply would always be, 'Maybe that's just your scrambled brains talking'. Therefore I can't accord objective validity to *any* hierarchy among my thoughts.
>
> But it is not possible to argue in this way, because it is an instance of the sort of argument it purports to undermine . . . There just isn't room for scepticism about basic logic, because there is no place to stand where we can formulate or think it without immediately contradicting ourselves by relying on it. (Nagel 1997: 62)

[2] Its being so is an indication of how distant this understanding of transcendental argument is from Kant's: as Kant intended it, the Second Analogy cannot stand alone, as an *ad hominem* refutation of Hume.

[3] (Nagel 1997); all further references to Nagel are to this book.

Although Nagel probably would not welcome the description,[4] there is no real reason not to count this a transcendental argument on the understanding outlined.

One might hesitate on account of its talk of 'immediate' self-contradiction. Transcendental arguments are conventionally compressed into the form of a *modus ponens*, whose conditional premise has the sceptically undeniable starting point as antecedent and the intended conclusion as consequent. The initial distance between these is then a measure of the ambition of the argument, ambitiousness being reckoned characteristic of the breed. In Nagel's argument the initial distance is slight, with the conditional premise verging on tautology. But this cannot disqualify the argument. Initial distance is at best a subjective measure—if the argument is good, it was only ever an apparent distance—and hardly suggests a criterion.

One might instead think that to include Nagel's argument would undercut a shared sense of something genuinely distinctive in transcendental arguments. To an extent that is true, but not my fault. It is just a consequence of trying to extract an understanding of 'transcendental' from that of 'transcendental argument', where those arguments are characterized in turn by an extrinsic feature of their conclusions—whether philosophical sceptics have chosen to doubt them.

To an extent, though, the complaint is false, at least if we restrict attention to arguments that are self-contained in the way Nagel's argument clearly is. Conclusions endorsed through such arguments are shown to have a distinctive (Nagel says 'dominant') 'position in my system of beliefs' (p. 65). They are presupposed in any enquiry into their truth, and for that reason rationally inescapable: 'we cannot get outside them'; 'that is why they are exempt from scepticism' (p. 64). This exemption is their strength.

But it is surely also very natural to regard it as their weakness. Being exempted from scepticism seems somehow less creditable than proving invulnerable to it, much as being exempted from an exam seems less creditable than passing it. If it is genuinely not 'possible to scrutinize these thoughts without presupposing them' (p. 64), then we cannot regard

[4] In *The Last Word* 'transcendental' is typically part of another phrase, 'transcendental idealism', denoting a position Nagel rejects.

them as having withstood any kind of scrutiny, or as having been vindicated through it. Their 'position in my system of beliefs' establishes them as measures that cannot themselves be measured. From that it can seem to follow that nothing other than their position in the system can contribute to their fixity: that they are held fast—the thought runs—*only* by the movement that surrounds them.

3. Three dimensions of logic's generality

Maybe that is how it is. It may not yet be worrying. As a way of asking whether it should be, we can ask, how much of what Frege's characterization of logical laws led us to hope for can be delivered by an argument like the one just borrowed from Nagel.[5]

The quotation I gave from Frege in §1 includes three signs of generality: logical laws hold 'with *the utmost generality* for *all* thinking, *whatever* its subject matter'. Whether or not Frege exactly intended this, we can take them to indicate three distinct dimensions in which logic is general.

First, and most simply, logical laws hold 'with the utmost generality', or (in Kant's phrase) with 'strict universality'. They brook no exceptions, not even hypothetical or merely possible exceptions (Frege 1884: §14), and so are *a priori* (Kant 1781/7: B4).This much can be accounted for by Nagel's argument, through the idea that basic logical laws are implicit in the terms of enquiry, or in our grasp of the concepts employed in it. This idea has a place in Nagel's discussion. 'We cannot conceive of anyone's positively believing [that a simple logical law is false],' he says, 'because we cannot attribute both understanding of and disbelief in it to the same person' (p. 64). But it can only have a small place. Perhaps it is true that *conjunction* is that concept to possess which one must acknowledge (as primitively compelling, and so forth) the laws of conjunction introduction and elimination. But it seems no more true than that *fortnight* is the concept governed by the schema, 'it lasted n fortnights iff it lasted $2n$ weeks'. So it can tell us very little about the distinctive status of logical laws. 'What use

[5] This circumlocution is meant to register that the comments in this and the following sections are directed only to the argument just presented. Without saying very much more than I have or can about the surrounding context from which this simple argument has been ripped it would be absurd to regard them as an evaluation of Nagel's broader stance.

can the all-embracing world-mirroring logic have for such special twiddles and manipulations?' (Wittgenstein 1922: 5.511).

A first advance on this is marked by Frege's claim that logical laws hold 'whatever [the] subject matter': these laws are completely general in their application; the concepts they involve are topic-neutral. There are weaker and stronger understandings of topic-neutrality. The weaker, negative construal holds only that logical laws are not confined by their content to any specific domain of application or enquiry. The stronger, positive construal holds that these laws are in force in every enquiry. The negative construal is compatible with the weaker thesis that, through any enquiry or evaluation, some laws stand fixed. The positive construal is needed to hold that some laws stand fixed through every evaluation. In a brief comment on the image of Neurath's boat Nagel is clear about this distinction, and clear that he intends the stronger construal (p. 65). This stronger construal depends, I think, on a commitment to the unity of logical laws, such as Frege displayed in holding that 'we have only to try denying any one of them and complete confusion ensues' (Frege 1884: §14). Nagel shares this commitment. Interestingly[6] he is drawn towards an account of it along Tractarian lines: 'Certain forms of thought . . . force themselves into every attempt to think about anything. Every hypothesis is a hypothesis about how things are and comes with logic built into it' (Nagel 1997: 62). All of logic, it seems, is carried in the general form: *this is how things are* (cf. Wittgenstein 1922: 4.5). Without some such account even (weakly) topic-neutral concepts would still be 'special twiddles'.[7]

Frege's observation that the laws of truth have authority for '*all thinking*' introduces—or at least seems to introduce—a third dimension of generality, on which even a unified system of laws (and the concepts they involve) might seem to be no more than one special location (and the concepts, in still another sense, just 'special twiddles'). The location is special, to us, because it is ours. But is our way of thinking, and the laws whose acknowledgement essentially structure it, the only way there could be, or the only system of laws that could play that distinctive structuring role?

[6] This is interesting, I think, because this early-Wittgensteinian thought is of a piece with ideas of Frege's sketched in §8 below, and whose role in countering a relativism inspired by the later Wittgenstein is the topic of §§6-9.

[7] The traditional 'transcendentals', *good*, *beautiful*, *true*, and so on, were regarded as weakly topic-neutral. Even accepting this we can still ask, for any list of them, why are there just *those*?

If there is a real question here, then it is hard to see how anything in the style of argument we have borrowed from Nagel could begin to address it. More than that, the strategy of the argument encourages us to think that there must be a real question here. The status this argument assigns to logic is one attaching to a distinctive position 'in my [our] system of belief', a position that is identified only internally, through its presuppositional relations to other elements of this system. The unease sown by this internal identification is nurtured by the fact that Nagel's core analogy for the inescapability of logic is the inescapability of the *cogito* (which, as all first-years learn, limits what I can coherently think, not what might be so). A first-person formulation is equally important to many of his leading statements of it: 'I can't regard it as a possibility that my brains are being scrambled, because I can't regard it as a possibility that I am not thinking' (Nagel 1997: 62); 'we cannot leave the object-language [= our language] behind, even temporarily' (p. 58); and again, 'we cannot get outside' logic (p. 64). This accumulation of essentially located idioms carries a strong invitation to ask how what they describe would present itself to a less embedded standpoint. And the image that question inevitably conjures up is of logic as at the centre of gravity of the body of our thought: if the body moves, so will its centre. It may well be that we cannot stabilize that image, just as we cannot measure where, along the third dimension of generality suggested in Frege's remark, our measure-of-all-things falls. But that is exactly what the image itself would predict.

4. Metaphysical abstinence

Of course, that this image is encouraged by the style of argument borrowed from Nagel is nothing like enough to make it troubling. In presenting it as such I am simply presuming on what Mark Sacks, in his discussion of these matters in *Objectivity and Insight*,[8] summarizes as a 'post-metaphysical orientation'. It is characteristic of this orientation, for instance, that when Frege says that the laws of truth 'are boundary stones set in an eternal foundation, which our thought can overflow, but never displace' (Frege 1893: xvi), we can hear that only as a rhetorically apt expression of their fixity, not as any kind of explanation of it. By contrast, when

[8] (Sacks 2000); except where indicated further references to Sacks are to this book.

Descartes held that certain immutable natures, and the principles that expound them, were fixed by God's decree, he intended a genuine explanation; similarly, though in a way oppositely, when Kant grounded those same principles in the essential structures and activities of a discursive understanding, as delineated in his transcendental psychology. A post-metaphysical orientation is one that has withdrawn from such claims. It will countenance no attempt to stabilize basic norms by anchoring them in brutely given metaphysical structures, whether of the mind-independent world or of the mind itself.

I can just presuppose that orientation here for two main reasons. The first is that Sacks has already done the work of elaborating and motivating it, in Parts I and II of *Objectivity and Insight*. Through the 'selective historical survey' presented in Part I he shows how viable alternatives to Kant's model of the mind (as synthetically structuring experience out of an atomized 'spray' of data for intuition) would undermine his attempt to demonstrate the interdependence of the unity of the mind and that of an objective world (Sacks, 2000 141). In Part II, under the heading of 'world-driven scepticism', he explains how various movements in twentieth-century thought have surrendered the conception of a normatively autonomous subject, converging in consequence on 'a naturalized, intersubjective model' which threatens to leave 'the voice of reason . . . tamed, parochialized' (p. 163). In this account Sacks puts a lot of weight on some arguments that I think cannot bear any (for example, the rule-following argument as construed by Kripke, p. 156), and draws conclusions from others that I am at least unsure of (for example, that there is some good sense in which, 'given [semantic] externalism, I cannot strictly speaking claim to know what my beliefs are', p. 159). But the second reason for brevity here allows me to skip over such quibbling. Sacks brings out clearly that the challenge to universal norms depends on no more than 'metaphysical abstinence' (p. 217)—a distrust of 'unexplained and possibly inexplicable transcendent grounds' (p. 216)—together with the undeniably mundane core of a basic 'epistemic duality' (pp. 187–91), that some of what we think about things we think because of the way we are rather than how they are. The second is enough to prompt the question, whether the basic norms we acknowledge can reach further than the systems of thought within which they are 'presuppositionally nested' (p. 191); the first is enough to refuse any reassuring answer. That refusal is enough. I do not need to go as far as Sacks does towards showing

us the river-bed shifting. It is enough for my purposes that we can lose our grip—or seem, temporarily, to lose our grip—on why it should not.

Sacks summarizes the options from here through his distinction of transcendental *features* and transcendental *constraints*.

> Roughly, a *transcendental constraint* indicates a dependence of empirical possibilities on a non-empirical structure, say, the structure of anything that can count as a mind. Such constraints will determine non-empirical limits of possible forms of experience . . . A merely *transcendental feature*, on the other hand, is significantly weaker. Transcendental features indicate the limitations implicitly determined by a range of available practices: a range comprising all those practices to which further alternatives cannot be made intelligible to those engaged in them. (Sacks 2000: 213)

In these terms the weakness of the easy answer lies in a mismatch between its aspirations and its methods: it hovers between 'two quite distinct positions: one which allows transcendental constraints, by still appealing to a metaphysical order, and one which prescinds from metaphysical commitments, and affords no more than transcendental features'. The argument claims the advantages of the first, promising 'standards of evaluation that are not merely contingent and cultural'; but its self-contained character signals its reluctance to make the kind of 'appeal . . . to . . . transcendent grounds' needed to sustain that promise.[9] Its internalist methodology is suited 'to the recognition only of transcendental features'. For all it shows, 'we are . . . left with the possibility of unconstrained empirical relativism . . . All standards of evaluation are . . . ultimately indexed to a given set of practices, to the tradition in which we participate'.[10]

5. Tests of strength

This may well seem an odd conclusion. After all, this kind of 'reductionism', or 'a subjectivist reinterpretation of reason', was precisely the kind of 'scepticism' about logic that Nagel's argument was designed to defeat (Nagel 1997: 9). The problem, I think, lies just there: in the ambition to *defeat* this scepticism, rather than to understand and disarm it.

[9] Here it is specially relevant to remember the warning of §1, that the easy answer is merely an extract from Nagel, and not a representation of his overall position.

[10] All quotations in this paragraph are from (Sacks 2000: 216).

The problem stands out most clearly in another of Nagel's comments on Descartes' willingness to contemplate there being alternatives to eternal truths. Descartes wrote:

> Again, there is no need to ask how God could have brought it about from eternity that it was not true that twice four make eight, and so on; for I admit this is unintelligible to us. Yet on the other hand I do understand, quite correctly, that there cannot be any class of entity that does not depend on God; I also understand that it would have been easy for God to ordain certain things such that we men cannot understand the possibility of their being otherwise than they are. And therefore it would be irrational to doubt what we do understand correctly just because there is something which we do not understand and which, so far as we can see, there is no reason why we should understand. (*Objections and Replies* VI.8, as quoted by Nagel 1997: 60)

Nagel comments:

> This implies a hierarchy among *a priori* judgements that is unpersuasive. The idea is that if we believe *G*, and *G* provides an explanation of why *I* would seem to us inconceivable even if it really wasn't, then it is reasonable to regard *I* as possible though we cannot conceive how. This makes sense as a general account . . . The trouble is that in this case, the inconceivability of *I* is so unshakeable that (by contraposition) it undermines confidence in *G*: It is impossible to believe that God is responsible for the truths of arithmetic if that implies that it could have been false that twice four is eight . . . Structurally, this argument of Descartes is precisely the same as is offered by those who want to ground logic in psychology or forms of life, and the same thing is wrong with it. (Nagel 1997: 60–1)

If the cases are structurally identical, then so, presumably, should be our responses to them. Suppose that Nagel is right in *both* cases about the strength of the opposed convictions. What are the results? In Descartes' case we might (too?) readily agree that arithmetic trumps (an already contested point in) theology, and that the right response is simply to abandon the conviction that God is omnipotent in any sense requiring arithmetic to be subject to His will. But Nagel can hardly intend us to argue in parallel with this—simply 'by contraposition'—that we, our faculties, and our ways of thinking are *not*, after all, contingently shaped and culturally embedded. That belongs so much to our framework that we would not know *how* to give it up.

Presenting matters as a trial of strength, as Nagel does here, cannot be the right approach to the sources of relativism, because it gives us no idea of *what to do* with the vanquished force. This is simply the application to

the case in hand of a point often made about scepticism in general, that it is unhelpful to conceive in adversarial terms of an issue in which victory for either side would be no better than a stand-off. I am not for a minute questioning that the *last* word needs to lie with basic rational norms that stand unindexed and unqualified. The problem, which Nagel has surely done as much as anyone to make vivid, is to understand how that is possible. Remarking that these basic thoughts 'dominate others' (p. 64), or that they 'force themselves to the top of the heap' (p. 63), hardly helps with that.

We need another approach. And, since I have followed Sacks into this bog, I want to see whether he can lead me out of it.

6. Sacks' broad strategy

From the literature on the prospects of transcendental argumentation Sacks highlights two key problems, both of them prominent in the work of Barry Stroud and responses to it.[11] The first is the issue of the *inference to reality*: it is the problem of how, or whether, arguments tracing how we cannot but think can yield conclusions concerning how things are. The second Sacks calls the problem of the *universality of inference*: it is the problem of explaining how principles that owe their standing to their location within a system of thought can claim any broader authority, of how, at the limit, they can claim authority 'for *all* thinking'.

To speak impressionistically, in the first place, it seems obvious that the two problems are connected. Both question how matters to do with the internal ordering of thought can reach outside that order—to the world, or to other ways of thinking about it.

In the second place, the connection was borne out in our initial response to the 'easy' answer. The problem of the inference to reality arose then with the thought that being exempted from scepticism is less creditable than withstanding it. An inescapable commitment of enquiry stands firm through the progress of enquiry, whatever direction it takes, and so is insensitive to the results of enquiry: it is insensitive, in other

[11] (Stroud 1968) is the crucial article; (Stern 1999a) usefully assembles a variety of responses to it.

words, to how things are.[12] But then how can we regard it as knowledge of how things are? (cf. Sacks 2000: 200, 299). At best, it seems that we should regard it as knowledge of how things will inevitably appear, insofar as this is determined solely by invariant aspects of the method of enquiry, and independently of its particular course. But that 'best' is available only on the presumption of something that secures genuine invariants in the method of enquiry. If we lose confidence in that, then this retreat from the inference to reality turns immediately into the problem of universality. (And a straightforward re-run of the inference-to-reality problem, applied this time to whatever is imagined to secure those invariants, stands ready, if need be, to shake that confidence; cf. p. 304.)

In the third place, the connection of the two problems is confirmed by a very simple argument. Since there is just the one world, to which all thinking is answerable, then if we could sustain the inference to reality, so as to hold that the conclusions of transcendental arguments delineate the order of that world, and not only of our thinking, we would *thereby* have resolved the universality problem. Our conclusions would have authority for all thinking, simply by virtue of its being answerable to the reality thus delineated.

This very simple argument not only connects the two problems, it suggests a priority between them. First, establish—somehow—an independent ontological standing for the structures our thinking is internally compelled to respect. Second, resolve the problem of universality through the straightforward thought that *all* thinking is then answerable to those same structures.

What I think is most interesting and important about Sacks' response to the two problems is that it reverses this suggested priority. The problem of universality, he contends, is 'of broader scope' (p. 273), and the one that we primarily need to attend to; if we can resolve the problem of universality, a solution to the inference-to-reality problem will drop out. This implies that the notion of answerability to something independently established can have nothing like the place in a solution to the universality

[12] 'Insensitive' is used here naïvely. There are those who hold that knowledge is 'sensitive' belief, and they no doubt have defensible things to say about how this applies to knowledge of necessities, where sensitivity, in their technical sense, seems to come for nothing. Whatever they say will quite properly start further along, reasonably presuming that necessities make a special case calling for special treatment.

problem that we just imagined. A solution must instead explain how it was a mistake to regard the structuring principles of thought as answerable to *anything*. Sacks' central suggestion is that this is mistaken because these principles are the very *source* of answerability. For thought to be so structured *is* for it to be answerable, and it makes no sense to suppose that in this respect it could be answerable to anything whatever.

In the terms already borrowed in §4, the promise of this suggestion is to deliver transcendental features which can claim unrestricted universality although they do not need—or rather, because they cannot intelligibly be thought to have—external constraint.

7. The three main planks of Sacks' construction

Sacks' working out of this proposal depends on three major elements. Because what is most relevant to my question is only the broad structure of the proposal, I will do little more than mention these elements, limiting comments to the most important of the working linkages between them.

7.1 The interdependence of subjectivity and objectivity

The first element is a sustained argument that even the most minimally heterogeneous experience, one allowing for any differentiation in its presented content, requires a unified objective backdrop (Ch. 7). This argument is in its early parts a replacement for, and in later parts a reconstruction of, Kant's transcendental proof of the interdependence of subjectivity and objectivity. Its starting point is more minimal yet than that of the Transcendental Deduction; its conclusion a more abstract version of that of the First Analogy. I will not attempt to give the flavour of this argument in a paragraph. For architectural purposes, the salient point is that, like Strawson's 'reconstructive approach' (Sacks 2000: 271), it aims to recover the force of Kant's insights without relying at any point on his transcendental psychology. The aim of Sacks' construction as a whole is to 'combin[e] . . . genuinely resilient universality and metaphysical abstinence' (p. 217). This argument, offered as the first plank in the construction, exemplifies the abstinence. Defence of its claim to universality rests on the other planks (pp. 271, 285).

7.2 *The normative matrix*

A credible claim to universality must accommodate realistic cultural and historical variation (p. 303). The second main element in Sacks' construction meets this need. The supposition that practices, world views, or forms of life might change requires that these set-ups 'be placed within a common system of coordinates'. Transcendental arguments will then properly be directed towards this 'common normative matrix', and not any of the particular, contingently conditioned loci within it.

The shape of this idea is appealing; its implementation, for my aims, less so.

> We might consider that logic as it is at one locus might not be the same as it is at another. There might indeed at different loci be different conceptions of logic between which there is nothing common that can be detected from the perspective of either of those loci. That it is *logic* that is being contrasted will, of course, not be detectable from those loci. Such relativism still makes sense . . . because for there to be meaningful disagreement between two loci, it is not necessary that the difference in question be detectable to the parties to the disagreement. All that is needed is, as we might put it, that there be some matrix in relation to which the normative practices of the parties can be calibrated. That is enough to render a possible disagreement between two loci that cannot be detected, or at least cannot be accurately delineated, from either. The background matrix will in the case envisaged contain a set of loci between which family resemblance holds, such that each can be recognized by its immediate neighbours as engaging in logic, but given the intransitivity of familiarity (or of translation), we can still allow that there might be loci so unfamiliar to each other that that recognition fails to hold between them. We can still say, however, that although what logic seems to consist of essentially from within any one locus might be no more than the shadow of necessity cast by the familiar practices that shape the horizons in that vicinity, nevertheless there is, underlying that parochial identification to which there are alternatives, a single conception that is common to all loci, and to which there can be no alternatives. This is the conception comprising possibly nothing more than the disjunctive specification of the conceptions of logic at the relevant family of loci. There can be no alternatives to this framework, since any candidate alternative is simply taken up as another possibility within it (as another disjunct, as we might put it). (Sacks 2000: 290–1)

I'm afraid I can make little of the last suggestion here. A disjunction (union) of logics is not a logic, however minimal. An intersection might be. But Sacks' concern to acknowledge the non-transitivity of familiarity—the idea that, by wandering from cognitive neighbourhood to neighbourhood, one might wind up somewhere impossibly foreign—strongly suggests that the intersection would be empty. Local accessibility is surely too weak to secure 'a single conception that is common to all loci'. Nor, I think, does the

notion of family resemblance help to make available the needed 'single conception'. We could grant (though I am myself sceptical even about this) that sensitivity to overlapping resemblances might sustain a conception *of* the space of variation, as it were, an overview from outside of it of what holds the space together; but such an overview stands at the wrong level to be the conclusion towards which transcendental arguments are directed. What Sacks' overall construction demands from this second plank is not a conception *of* a space, but rather something that simply articulates a conception one has by virtue of occupying any locus within that space.[13]

A second way of raising effectively the same reservation would be to say that the need to accommodate variation so far gives us only a matrix of norms, not a normative matrix: we have the first when the points represent particular normative outlooks; for the second the points must stand in normative relations. That requirement stands in some tension, I think, with the suggestion at the end of the quoted paragraph, that the matrix will, just by virtue of the job it was introduced to do, take up any alternative as 'another possibility within it': a space so accommodating has no geometry.

In one respect these reservations are entered too soon. This second element in Sacks' construction is not intended to stand alone. The final element is needed to explain how the normative matrix can meet the description given of it, and play its role as the proper aim of genuinely

[13] We will be better placed after the comparisons with Frege sketched in the following section to recognize a further problem for this idea of a family resemblance conception *of* the space of variation. Such a conception would be, in effect, a conception of a role that is variously fulfilled by whatever counts as logic in each of the particular outlooks, something that supplies an '*it*' for the thought that 'these are all different ways of doing *it*'. But it is essential to Sacks' overall construction, at least as I understand it, to maintain that we do not have a notion of some goal that various cognitive outlooks in their different ways achieve that could be prior to or independent of our sharing in whatever conception is common to all outlooks. The relevant goal, or the '*it*' in question, is thinking about the world. And the idea behind the thought of 'different ways of doing *it*' is that, to make its 'vertical' connection with the world, thinking must be regulated or disciplined by 'lateral' connections, such as are supplied by 'a logic'. This use of the idea, however, subverts the priority implicit in the Kantian thought on which it draws (1781/7: A104–5). Or, to put the same point in Fregean terms, one cannot presume on the notion of thought, and then explain logic as what regulates thought; instead, a thought is precisely something subject to logical laws. (I would make just the same complaint against the family resemblance notion of a proposition that Wittgenstein offers us in the *Investigations* (1953: §§134–6), in explicit rejection of Fregean elements in his early thought.)

universal transcendental argument. Sacks makes this point by saying that, while the matrix can be made by definition all-inclusive, it cannot simultaneously be made by definition unchanging—at least, not so long as we allow that all normative outlooks, even the most inclusive, are subject to 'the sheer force of biological change' (p. 293). And a changing matrix, in which loci can come and go, is effectively equivalent to one with 'no geometry'.

Even so, the eventual role of the matrix demands, I think, a more thorough connectedness than the current picture provides, and I do not see why we shouldn't have it. In part this may be because (as mentioned in §4) I am not persuaded by some of the particular arguments for culture-dependence that Sacks invokes. But there is also a structural point. In the final scheme, the normative matrix is to accommodate realistic cultural-historical variation. It need not be designed to accommodate everything that merely *seemed* to be possible while the threat of unconstrained relativism was in force. Mutually unrecognizable logics are (I hope) a nightmarish fantasy belonging to an earlier stage of the argument.

How does this second element of the construction connect with the first? The normative matrix supplies 'an elusive necessary condition for the universality of transcendental arguments' (p. 292), something towards which such arguments might coherently be directed. We will clearly want more than a necessary condition—more than merely the existence of a target and the bare possibility of a hit. Again, that will emerge only in the finished construction.

7.3 *The dissolution of fictional force*

By a 'fictional force' is meant 'a propulsion to belief that does not survive critical reflection on its evidential base' (p. 297). We have seen twice now (first in §2, and then again in the sketch given in §6) how the problem of the inference to reality, and the relativist worries it brings in tow, can arise from a critical reflection of that kind. This reflection starts out from the thought that certain basic commitments are so centrally a part of how any question is addressed that we cannot but think in accordance with them; and it turns that into the question, what assurance we could have that thinking of things in accordance with these commitments amounts to thinking of things as they are.

If we follow the train of reflection this far, then it looks as though we can neither go back nor stay where we are. The problems of going back

were considered in §5. The problem of staying where we are is that the position we have reached is, by its own standards, very obviously unstable. This position encourages us to look askance at the basic commitments it concerns, on the grounds that their role as the evaluators of evidence screens them from any possible counter-evidence. But in the process the position has drawn exactly the same kind of protective screen around itself: it has made plain in advance that any consideration offered to lift the shadow it casts over our basic commitments can amount to no more than yet another manifestation of them. So, if neither sticking nor retreating will do, the remaining option is to press the critical train of thought further. And, in the challenge it has so far led us to, there is just one salient object for further criticism: the minimal presumption this relativist challenge shares with its absolutist target, that there is a way things are. That there is is indeed something we cannot but think—it is another inescapable commitment of enquiry, and perhaps the most basic one. But as such it ought in consistency to be within the scope of the critical question, rather than something presumed in framing it. Without this presumption, though, it seems that the question cannot be framed at all.

That is one way of motivating Sacks' suggestion that 'the antidote' to relativist worries should be looked for 'in the poison itself' (p. 297 n.41), but it is somewhat different from the route he takes.

Sacks' own route starts from 'the epistemic duality between the uncritical and the critical reception of the order of things'. This duality opens up the possibility that an order of things which ordinarily we 'uncritically read off from the world', and accept as 'an order imposed on us by the world', might be critically recognized as instead an order 'that we ourselves, behind our cognitive backs, have imposed on the world'. Relativism thrives so long as this duality is conceived as applying only to the particular structurings of the world supposed to reflect our various practices. But the thought of this duality can be extended, and applied 'not merely to the order of things, but to the very notion that there is some ontological base to be ordered' (p. 297). The transcendental argument that is the first element of Sacks' construction then suggests that the thought *should* be so extended. That argument concluded that 'the assumption of an objective domain' (p. 299), antecedent to and imposing upon experience, is itself a basic commitment of experience. So, in critical or reflective mode, we should count it no more than that: ordinary acceptance that there simply is a way things are 'does not survive the critical insight that the

structure of experience is such as to compel confrontation with an objective order at the empirical level regardless' (p. 299). Sceptical or relativist worries then appear merely as a symptom of not pressing the critical thought to its proper extent, or of exposing only some of the 'fictional force' attaching to our basic commitments to its corrosive effects. These worries still trade in the notion of a given 'ontological base, the vagaries and mutations of which . . . stand to render unstable the apparent fixity of our normative structures'; but that notion is now 'recognized to be one that should not be brought into play in a properly critical enquiry' (p. 301).

It is essential to this resolution that its critical retreat from the assumption of a way things simply are is neither the nihilistic denial that there is such a way, nor the idealist claim that the appearance of one is a projection of mind. Both of those are themselves ontological claims (p. 278), and so in this context are properly counted as unusual versions of realism. The claim is instead this:

> that the appearance of an ontological base given antecedent to experience and brutely imposing upon it is itself critically recognized to be a construal that is in accordance with an imposed form; a form imposed not by *us* but by the very structure of experience. (Sacks 2000: 299)

The three elements of Sacks' construction come together in this claim. That to which transcendental argument is properly directed is precisely that universal structure, represented in the normative matrix, which first introduces, or 'sets up', an ontological conception of objectivity. This structure can be grounded neither in the world, nor in the mind, since it is prior to both: it is the structure that at once opens up the world to us, and defines mind as standing in coordination with that world. So there is 'no space for the worry' that this structure might be shaped or altered by 'influence external to it' (p. 301). To put the point incautiously, nothing is external to it.

8. Two echoes

Before raising some questions about this construction, I will mention in this section some ways in which it echoes earlier thoughts. Because the aim here is not to advance the case, but just to indicate why it might be attractive, I am content to be sketchy and impressionistic.

8.1 Kant

The first echo is too obvious to labour. Sacks' construction promises a finality that the 'easy' answer could not, precisely because it reintroduces aspects of the meaning of 'transcendental' that have dropped out of the common understanding of it described in §2:

> I entitle *transcendental* all knowledge which is occupied not so much with objects as with the mode of our knowledge of objects in so far as this mode of knowledge is to be possible a priori. (Kant 1781/7: A11–12/B25)

Effectively, the common understanding sketched simply omits the last clause of this explanation, and says that things transcendental have to do with our mode of knowledge rather than with its objects, or more specifically, with commitments presupposed by or implicit in our ways of knowing, not with the things we know about. §3 briefly mentioned ways in which that kind of idea has been used to explain something's *a priori* status. But the transcendental is a narrower category. Sacks' proposal puts back the last clause. According to it, something structurally embedded in a mode of knowledge will be transcendental only if it has an essential role in explaining *a priori* how a mode of knowledge so structured is knowledge of objects.[14]

8.2 Frege

The second echo may be less obvious, but it is important to me in indicating how logic might be transcendental in this more demanding sense.

'[T]he laws of logic,' Frege says, 'are nothing other than an unfolding of the content of the word "true". Anyone who has failed to grasp the meaning of this word . . . cannot attain to any clear idea of what the task of logic is' (Frege 1979: 3). One fruitful way of attaining that grasp, I think, is to reconsider Sacks' construction, and to imagine *the very structure of experience* as replaced throughout by *the structure of truth*.

The notion of truth is unique, or *sui generis* (1918–19: 354). It does not belong to the description of the world considered independently of thought: its laws are 'not laws of nature . . . but laws of the laws of nature' (1884: §87). Equally, it does not belong to the description of mind

[14] This is the central way in which Sacks' proposal differs from the widespread but less credible idea that, because networks of concepts establish ways for thought to be answerable to the world, they are not themselves answerable to the world.

considered independently of its answerability to the world it represents: that is the province of psychology, which has 'no inherent relation to truth whatsoever' (1979: 2).[15] Truth's sphere is instead *thought*—mind in its essential relation of answerability to the world (1918–19: 369).

The priority Frege assigns to the notion of truth implies that the understanding of this relation is prior to that of its relata, which are properly understood as abstractions from it. The understanding of mind aimed at in philosophy is an understanding of a subject rationally answerable to the world; its understanding of the world, conversely, is as the object of rationally structured thought. Thought, the sphere of logic's concern, is 'just what opens up the external world to us' (ibid.).

Again, the priority of the notion of truth implies that this region of concern is not one that we could somehow map antecedently, only subsequently recognizing logic's governance over it. Rather, a thought just is something for which the question of its truth can arise (1918–19: 353, 1979: 174), so something subject to the laws of truth. So we appreciate what a thought is, and what answerability amounts to, in acknowledging these laws. And that is why, as I said in summary of Sacks' proposal, it makes no sense to suppose that, in respect of its being structured by those laws, thought might be answerable to anything whatever.

The echoes here are, I'm sure, much more than coincidences. I hope they are enough to indicate how, in following Sacks' construction, I have, without much explicit mention of logic, been pursuing an answer to my question. At any rate, they explain my interest in asking, as the following section does, whether the construction goes through.

9. Two questions

9.1 A retreat from ontological commitment

At the end of §7 I said that Sacks' account, of how the structure that introduces answerability is invulnerable to external influence, could be summarized incautiously by saying that nothing is external to that structure. I first want to ask how incautious that was.

[15] Giving the laws of truth sway over a *third* realm was perhaps not the best way of making the last two points, but it was *a* way.

If we suppose, first, that it was not at all incautious, then the construction as a whole does go through. It certainly *sounds* incoherent to suggest that whatever determines that thought is confronted with a given order should itself be determined by that order. And it will clearly *be* incoherent if 'determined' is given any constructive or constitutive gloss: it cannot be that this basic normative structure, or the structure of experience, both creates and is a creature of the world it opens up. But that kind of gloss is hardly compatible with the 'metaphysical abstinence' Sacks' proposal aims to respect.

A different model of the incoherence, better suited to that abstinence, is offered in the way Sacks distinguishes his proposal from the idealist claim that the appearance of a way things simply are is 'merely a construct of ours, . . . something that we project onto the world' (Sacks 2000: 299). If we recognize this appearance as 'a construal in accordance with an imposed form', we cannot at the same time hold that this form is imposed by any way things (our minds) simply are. The incoherence of this would be analogous to that of ascribing responsibility for the mere appearance that *p* to the fact that *p*. It would involve simultaneously endorsing and bracketing commitment to there being a way things simply are.

For the most part this is the model that Sacks is content to work with in his discussion. It is one that allows that 'there could be . . . an ontological base', but insists that 'any assertions to that effect . . . [must] . . . appear blind to the very conditions of experience in which they are made' (p. 298). What awareness of these conditions calls for, then, is a suspension of, or a retreat from, 'ontological *commitment*' (p. 300, my emphasis). The model need not involve us in questioning the significance of assuming a given ontological order, but only in withholding endorsement from that assumption: 'there is no room to assert within a duly critical philosophical stance that there is a way the world is in itself' (p. 301).

It is by adhering to this model, I think, that Sacks is able to present 'the antidote' to relativism as already present 'in the poison itself', that is, as lying in an extension of the epistemic duality that generates it, where the duality is understood in much the same way that the relativist himself understands it.[16] Similarly, it is this model that informs the metaphor of 'fictional force', a propulsion to take appearances at face value. The metaphor suggests, and

[16] Hence, one way of questioning the model is to look at the paragraph where Sacks has to explain that 'the epistemic duality is *not* in play . . . in quite the same way as before' (p. 299, my emphasis).

the model implies, that the fictions themselves remain in place: critical reflection nullifies only the force that ordinarily attaches to them.

But I do not see how this model can be adequate to Sacks' purposes. The challenge to universality arises from the reflection that what seem to us to be necessities of thought, and are in fact inescapable commitments of our thinking, might be shaped by a reality in which they are contingently embedded. On the face of it this need not involve any assertion that there *is* an embedding reality: it seems to depend only on the supposition that there *might be*. To counter it, or to show that 'there is no longer the space for it', it must be shown that even this supposition is, as Sacks puts it, no longer 'acceptable currency' (p. 300), and that 'the [very] notion that there might be some such ontological base . . . [cannot] . . . be brought into play within a properly critical enquiry' (p. 301). This is more than a retreat from ontological commitment. It has to be instead a retreat from uncritical acceptance of the terms of ontological debate.

This could come as no surprise to Sacks. It is exactly what he argued in his first book, *The World We Found* (1989). In something close to its terms, the point we need can be put by saying that, since the conception of an objective world confronting enquiry is a requirement of the structure of experience, our understanding of what it is for there to be an objective world must be an understanding of it as an abstraction from the structure of experience (1989: 167).[17] It is just that understanding that would be distorted by,[18] and so leaves no space for, even the supposition of a world existing as an antecedent determinant of the structure of experience.

I have recalled this theme from Sacks' first book, not in the hope of elaborating or properly defending it, but as a corrective to misunderstandings allowed by the model described earlier in this section, according to which reflection brings about only a withdrawal of commitment. The first of these misunderstandings would have it that the conception of a way things simply are in which the relativist seeks to ground the structure of

[17] If this sounds idealist, try translating it, as the previous section recommended, into 'Fregean'. Prominent in the rendering might be, that the world is the totality of facts, and a fact is a thought that is true.

[18] As an illustration of the distortion involved, consider the effects of casting the natural world, to which empirical thought is answerable, in the role of antecedent determinant of the structure of empirical thought (the role previously occupied by Descartes' God and by Kant's transcendental psychology): the effect is to transform naturalism from the default stance of a 'post-metaphysical outlook' into a variety of dogmatic metaphysics—and surely the least appealing variety of dogmatic metaphysics yet invented.

thought is the same conception as is exercised in thought so structured. It is not. The first is an empty distortion of the second. But then I think it is hard to resist the conclusion that the notion of 'fictional force' is also a misunderstanding. On the one hand, the empty and distorted notion of a way things are *in themselves* does not get so far as to be a coherent fiction. On the other, there is nothing fictional about the force attaching to the notion it distorts, that of a way things are to which thought, in virtue of its essential structure, is answerable.

The upshot is that the summary of §7, that nothing is external to this structure, was only slightly incautious. A cautious statement might be that no understanding of 'something' makes 'something is external to this structure' both intelligible and true. (But, as always in these connections, it is hard not to think that the quotation marks get in the way.)

9.2 Circularity

Probably the most common reaction to Sacks' treatment of the universality problem will be to suspect that the whole construction must be circular. Its aim is to show that the conclusion of a transcendental argument can legitimately claim universal authority, one not indexed to the framework of thought within which it is developed and whose presuppositions it explores. The crucial move in the construction is the move that distances us critically from the notion of a way things simply are, encouraging us to recognize that instead as 'a construal in accordance with an imposed form[,] a form imposed . . . by the very structure of experience' (Sacks 2000: 299). But surely what is meant here by 'the very structure of experience' is the *necessary* structure of *any* experience. The phrase embodies a claim to universality of just the kind that is in question. But then how can we endorse this claim in advance of, and as a means to, the overall conclusion? It seems that we must accept this one transcendental argument as yielding a universally authoritative conclusion before we have shown that any such argument can do so.

The complaint is misplaced. It is true that Sacks' overall conclusion (call it *C*) depends on the lemma (call it *L*) established by the transcendental argument which he presents as the first element of his construction. It is also true that *L* is a claim about *any* form of experience. It follows that the case for *C* will fail unless *L* is vindicated, and, since this is what it is, vindicated *as* a claim about *any* experience. The vindication of *L* lies, however, not in *C*, but precisely in the argument presented for it. What *C* supports is not *L* itself,

but rather a comment on *L*'s status, namely, that the unrestricted generality it claims for itself need not be compromised by any further indexing.

Even so, I think we should admit that moves of the kind just made are sometimes unconvincing.[19] Proof of a proposition with a certain status, and proof that it has that status, are, while formally separate, sometimes hardly separable in point of conviction: unless we can recognize how the (initial) proof confers that status on the proposition, the proof itself will remain in doubt. So it is important to understand why the relativist cannot fault the proof of *L*, and then the overall case for *C*, on those grounds.

Here the starting point is that the relativist and his opponent are *agreed* that the *apparent* content of *L*, the lemma proved by the transcendental argument, is that of a universal generalization embracing *any* form of experience. The relativist, unable to understand how any argument could yield that much, contends that *L*'s apparent content cannot be its real content. We must recognize, he claims, a hidden parameter or index, whose effect is to turn '*any* experience' into 'any experience that *we* can conceive of'. And he then asks how we could warrant generalizing universally along the further dimension introduced by this additional parameter, a generalization that would recover the intended force of *L* in the new form, 'any experience that *anyone* could conceive of'. This account of the scenario implies, first, that the relativist cannot deny the real separation between, on the one hand, the proof of *L* itself, which is prior to *C* and, according to him, addresses only the first dimension of generality, and on the other, the claim about the status of *L*, which derives from *C*, and which *he* will conceive of as addressing that further dimension of generality. More importantly, it implies that the non-relativist need not share that conception of what *C* is intended to yield. It is only on the relativist's presumption, that the transcendental argument cannot prove exactly the unrestricted universal generalization it seems to prove, that we would have any reason in the first place to recognize the supposedly hidden index, or the further dimension of generality on which that index flags our particular location. The role of *C* is to deflect that presumption, rather than answering the question it generates. It is not a way

[19] The response that Bernard Williams offers to the problem of the Cartesian Circle in *Descartes* (1978: 200–4) is of just this kind, and would, I imagine, be generally regarded as unpersuasive.

of *making* the further generalization that the relativist envisages, but a way of making plain why no further generalization is needed.

Just the same can then be said—as indeed Sacks does say (p. 309)—about the inference to reality. What *C* provides in that case is again, not a way of making the inference, but a way of exposing the presumptions that seem to make the inference necessary.[20]

Typically insistence on the need for, and difficulty of, the inference to reality is taken to be a sign of robust (self-described) or uncritical (other-described) realism. So it is a further advantage of Sacks' construction that it explains how realism and relativism are fundamentally on the same side.[21]

10. The easy answer revisited

The conclusion just reached forces us to look again at the 'easy' answer to our question given in §2. When we first considered it I complained that it included nothing to address the final dimension of generality suggested in Frege's remark that the laws of logic are authoritative for *all* thinking (§3). What I have just claimed in §9 is that we should not look for an inference to span that final dimension of generality, but only a way of dissolving the pressures that seem to demand one. This suggests that my initial assessment of the easy answer was grossly unfair.

Maybe it was.[22] It depends, not on the argument that supplies the easy answer, but on the ambitions and attitudes that surround it. The initial

[20] Compare the response to Stern that occupies Sacks at pp. 274–85. Stern is *of course* right that, if the inference to reality required an 'external metajustification', a proof that our most basic 'internal standards of justification . . . are truth-conducive' (1999b: 59), it would be hopeless. The fantasy of a justification that would be 'external' in the required sense is one thing that could rightly be dismissed by Nagel's remark that 'in logic we cannot leave the object language behind, even temporarily' (p. 58). But then the image of ourselves as *having* to reach down from such an external perspective, to erase the index that our principles invisibly carry, is already clearly a fantasy.

[21] This conclusion rather neatly explains how Rorty, for instance, can be both at once.

[22] This is why I took care to talk about an argument extracted from Nagel, rather than 'Nagel's argument'. Whether any of the initial assessment applies to Nagel's own views depends on whether he shares realist presumptions of the kind described. Sacks interestingly suggests (p. 313 n.) that *The Last Word* contains hints of the kind of critical route to objectivity he would himself recommend. (One further indication of this, noted in §3, is Nagel's sympathy with the Fregean-Tractarian idea that logic is carried in the concept of truth, or in the concept of things' being thus and so.) I hope that Sacks' suggestion is right, because there is very much more in Nagel's book than this one easy argument that I would want to agree with.

assessment presumed on a robustly realist setting, one that wants the truths of logic, like every truth, to be answerable to the way things are in themselves. In that setting the argument fails, in the sense that it does not supply what the setting asks for. But that need not have been the fault of the argument itself.

Some might have had a sense right from the beginning that my initial assessment of the easy answer was slanted and unsympathetic.[23] For instance, when the answer proposed, as logic's particular strength, that it is 'exempt from scepticism', the assessment portrayed this as a weakness instead. That surely traded on an optional understanding of being exempted—one fed by an inappropriate analogy with someone who works the system to get out of taking an exam. Again, maybe so. My excuse is that this is the only understanding the robustly realist setting allows. Change the setting, and better analogies are available—for instance, that of a sovereign whose role in constituting the law makes it incoherent to call him to answer to it.

Change the setting, and we might as well say that the easy answer works. After all, it *ought* not to be hard to argue that to raise doubts about the standing of logic is simply wrong-headed. The harder part, for which I've very largely relied on Sacks, is to see one's way to accepting an easy answer.[24]

References

Bell, D. 1999. Transcendental Arguments and Non-Naturalistic Anti-Realism. In Stern 1999a: 189–210.

Frege, G. 1884. *Die Grundlagen der Arithmetik*. Translated by J. L. Austin as *The Foundations of Arithmetic*. Oxford: Blackwell, 1950.

—— 1893. *Grundgesetze der Arithmetik*, volume I. Part-translated by M. Furth as *The Basic Laws of Arithmetic*. Berkeley and Los Angeles: University of California Press, 1964.

[23] For instance, Antony Duff and Sandra Marshall straightaway had this sense when I presented an early draft of this material to a seminar in Stirling. I am grateful to them for suggesting the legal analogies mentioned later in this paragraph.

[24] Thanks to Adrian Moore and Mike Wheeler for reading and commenting on a draft. This paper was written for a meeting in 2008 of Mark Sacks' AHRC-funded project on Transcendental Philosophy and Naturalism. My thanks are thus due to Mark first for providing the context for this work, but then secondly, and in far greater measure, for providing its subject matter.

—— 1918–19. Der Gedanke. Translated as 'Thoughts', in B. McGuinness (ed.), *Collected Papers on Mathematics, Logic and Philosophy*. Oxford: Blackwell, 1984.

—— 1979. *Posthumous Writings*. Edited by H. Hermes, F. Kambartel, and F. Kaulbach. Translated by P. Long and R. White. Oxford: Blackwell.

Kant, I. 1781/7. *Kritik der reinen Vernunft*. Translated by N. Kemp Smith as *Critique of Pure Reason*. London: Macmillan, 1929.

Nagel, T. 1997. *The Last Word*. Oxford: Oxford University Press.

Sacks, M. 1989. *The World We Found: The Limits of Ontological Talk*. London: Duckworth.

—— 2000. *Objectivity and Insight*. Oxford: Clarendon Press.

Stern, R. Ed. 1999a. *Transcendental Arguments: Problems and Prospects*. Oxford: Clarendon Press.

—— 1999b. On Kant's Response to Hume: The Second Analogy as Transcendental Argument. In Stern 1999a: 47–66.

Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241–56.

Williams, B. 1978. *Descartes: The Project of Pure Enquiry*. London: Penguin.

Wittgenstein, L. 1922. *Tractatus Logico-Philosophicus*. Translated by D. F. Pears and B. F. McGuinness. London: Routledge, 1961.

—— 1953. *Philosophical Investigations*, 3rd edition. Translated by G. E. M. Anscombe. Malden, MA: Blackwell, 2001.

9

Strawson on Other Minds

Joel Smith

In the third chapter of *Individuals* Strawson offers a transcendental argument intended to disarm scepticism about other minds. The argument, frequently discussed, has had a mixed reception. Among recent commentators, Stern (2000) endorses a version of it, whereas Sacks (2005) argues that it fails. I too believe the argument to be unpersuasive, but I think that it is worth revisiting since, in my view, it fails for reasons not previously articulated.

In §1 I outline two versions of Strawson's argument, one ambitious and one modest. The ambitious argument fails for familiar reasons but, I argue, the modest version still has significant anti-sceptical potential. In §2 I introduce a number of criticisms of transcendental argumentation that call into question their claims to synthetic *a priority*, directedness towards objective reality, necessity, and universality. I argue that, modestly understood, Strawson's argument has some resources with which to respond to each of these criticisms. However, in §3 I point out that Strawson's argument confuses a number of importantly distinct questions and, as a result, fails to convince. I conclude, in §4, with some reflections on the prospects for a revised, naturalized version of Strawson's argument.

1. Strawson's argument

Before distinguishing the ambitious and modest readings of Strawson's argument, it is worth quoting him at length:

Clearly there is no sense in talking of identifiable individuals of a special type, a type, namely, such that they possess both M-predicates and P-predicates, unless there is in principle some way of telling, with regard to any individual of that type, and any P-predicate, whether that individual possesses that P-predicate. And, in the case of at least some P-predicates, the ways of telling must constitute in some sense logically adequate kinds of criteria for the ascription of the P-predicate. For suppose in no case did these ways of telling constitute logically adequate kinds of criteria. Then we should have to think of the relation between the ways of telling and what the P-predicate ascribes, or a part of what it ascribes, always in the following way: we should have to think of the ways of telling as *signs* of the presence, in the individual concerned, of this different thing, viz. the state of consciousness. But then we could only know that the way of telling was a sign of the presence of the different thing ascribed by the P-predicate, by the observation of the correlations between the two. But this observation we could each make only in one case, viz. our own . . . [But] what, now, does 'our own case' mean? There is no sense in the idea of ascribing states of consciousness to oneself, or at all, unless the ascriber already knows how to ascribe at least some states of consciousness to others. So he cannot argue in general 'from his own case' to conclusions about how to do this; for unless he already knows how to do this, he has no conception of *his own case*, or any *case*, i.e. any subject of experiences . . . the behaviour-criteria one goes on are not just signs of the presence of what is meant by the P-predicate, but are criteria of a logically adequate kind for the ascription of the P-predicate . . . [This conclusion] follows from a consideration of the conditions necessary for any ascription of states of consciousness to anything . . . But once the conclusion is accepted, the sceptical problem does not arise. (Strawson 1959: 105–6)[1]

1.1 The ambitious reading

A 'P-predicate' (short for 'person-predicate') is a predicate that if truly ascribable to an individual implies that that individual is conscious. An 'M-predicate' (short for 'material-predicate') is a predicate that could be truly ascribed to both persons and inanimate objects. Examples of the former are 'is in pain' and 'is playing football', examples of the latter are 'is six feet tall' and 'has fallen over'. I shall assume that associated with P-predicates are P-concepts, and that both denote P-properties. With this

[1] Also, 'When we take the self-ascriptive aspect of the use of some P-predicates, say "depressed", as primary, then a logical gap seems to open up between the criteria on the strength of which we say that another is depressed, and the actual state of being depressed . . . if the logical gap exists, then depressed behaviour, however much there is of it, is no more than a sign of depression. But it can only become a sign of depression because of an observed correlation between it and depression. But whose depression? Only mine, one is tempted to say. But if only mine, then not mine at all. The sceptical position customarily represents the crossing of the logical gap as at best a shaky inference. But the point is that not even the syntax of the premises of the inference exists, if the gap exists' (Strawson 1959: 109).

terminology in place, I suggest that we understand the central steps of the argument as follows:

1. Suppose that behaviour (that which can be observed) is 'a sign' of the presence of P-properties (which are not observable).
2. To be able to ascribe a P-predicate to another, on some occasion, I must know that such and such behaviour is correlated with so and so P-property.
3. But since we are asking how I can have come to be able to ascribe P-predicates to others at all, I must have learned of this correlation from my own case.
4. Learning from my own case depends upon the capacity to self-ascribe P-predicates.
5. But it is a necessary condition of the capacity to self-ascribe P-predicates that one possess the capacity to other-ascribe P-predicates.
6. So, the capacity for the self-ascription of P-predicates would be a condition of the capacity for other-ascription, which in turn is a condition of the capacity for self-ascription.
7. The capacity for self-ascription could not be a condition of itself.
8. So, behaviour is not a sign of P-properties but constitutes 'logically adequate criteria' for the ascription of P-predicates.
9. So, the sceptical problem does not arise.

This argument needs clarifying in a number of ways. Top of the list is accounting for what Strawson means by both 'sign' and 'logically adequate criteria'. Clarifying one of these terms ought to clarify the other since, as should be clear from step 8, Strawson takes them to exhaust the options. So let us take 'sign'. Signs, and the things of which they are signs, are distinct existences. Thus it seems reasonable to suppose that if behaviour is a sign then, presented with observable behaviour, my attribution to another of some P-property would have to be based on an inference (with 'inference' taken in some suitably broad way). This inference, Strawson tells us, would have to be based on knowledge of a correlation between behaviour type and P-property. The inference, then, is inductive rather than deductive.

So much for signs, what of 'logically adequate criteria'? On the assumption that the argument is valid, a logically adequate criterion ought to be nothing other than a way of knowing that is not an inductive inference. If 'logically adequate criteria' had some richer meaning, further justification

would be required for step 8. Quite plausibly, the category of non-inductive ways of knowing is exhausted by deductive inferences and non-inferential ways of knowing. The latter, in turn, might either involve the direct observation that the other instantiates a P-property, or the sort of non-inferential, defeasible criteria, familiar from (Hacker 1972) and (Wright 1980). Now, it seems highly implausible to suggest that one can deductively infer the presence of a P-property from the presence of some 'mere' behaviour. And I don't think that there is any reason to suppose that Strawson believed otherwise. Between the two non-inferential options, however, it is not so easy to choose.[2] There is some evidence that Strawson himself endorsed the direct observation view. He writes, 'X's depression is something, one and the same thing, which is felt, but not observed, by X, and observed, but not felt by others than X' (Strawson 1959: 109). But whatever specific view Strawson has in mind, it seems that the most that he can say, given the argument presently under discussion, is that our ways of telling that others are in certain mental states must be something other than inductive inferences from 'mere' behaviour.

It may be argued against this, however, that Strawson does in fact employ a richer notion of criteria; one that is unacceptably verificationist. To justify this, one might point to the opening sentence in the lengthy quotation above which appears to link the *sense* of talking about persons with ways of telling that they possess P-properties. Furthermore, it would seem that, in order to justify the argument's move from 8 to 9, criteria must be such as to satisfy two conditions. First, it must be that if one is in possession of the criteria for there being other minds, one thereby knows that there are other minds. Second, we must know that we actually are in possession of such criteria.

That the first condition must hold is obvious from the fact that, on a natural understanding, what the sceptic denies is that we know that there are other minds. So, surely, if the sceptical problem does not arise, this must be because we do know that there are other minds. The second condition is required since the fact that something is a criterion for there being other minds is not, in itself, sufficient to show that we know that there are other minds. We need to be in possession of the criterion and, if we can rely on this in the argument, we must know this.

[2] If, that is, the idea of non-inferential, defeasible criteria is coherent. See (McDowell 1982).

Now, suppose that to be in possession of criteria of a logically adequate kind for an attribution of some P-property to another is to be directly aware that they are in that state. This, plausibly, is one non-verificationist way to satisfy the first condition.[3] However, a sceptic might question the claim that we know that we are in possession of such a criterion. That is, the sceptic may argue that we cannot know whether we have actually experienced that another possesses some P-property, rather than it merely seeming to us that we have had such an experience. Thus, the sceptic can question our right to assert that the second condition is satisfied. Suppose, on the other hand, that the satisfaction of the second condition is not open to doubt by the other minds sceptic. Then it would seem that behaviour is being understood in a way that is common between those cases in which it is really expressive of mentality and those cases in which it is not. It is 'mere' behaviour. If so, how can our possessing such a criterion be sufficient for knowledge? The consequence is that it is hard to see how both of the conditions can be satisfied.

As is well known, Stroud (1968) claimed that Strawson's argument depends on a form of verificationism:

> Strawson's characterization of the skeptic is correct only if my possession of 'logically adequate criteria' for the other-ascription of a particular psychological state implies that it is possible for me to know certain conditions to be fulfilled, the fulfillment of which logically implies either that some particular person other than myself is in that state or that he is not. This must be a suppressed premise of Strawson's argument or an explanation of 'logically adequate criteria' . . . the skeptic is seen as maintaining both that (i) a particular class of propositions makes sense and that (ii) we can never know whether or not any of them are true. For Strawson the falsity of (ii) is a necessary condition for the truth of (i), and the truth of (i) is in turn required for the skeptic's claim itself to make sense. Therefore the success of Strawson's attack on both forms of skepticism depends on the truth of some version of what I have called the 'verification principle'. (Stroud 1968: 248)

Stroud sees Strawson as arguing that our being able know whether or not some class of other-ascriptions of P-predicates is true is a necessary condition of their making sense. But this is to tie the meaning of a class of sentences to our being able to ascertain their truth-value and this is a form of verificationism. Without the verificationism, the most that could be

[3] See (McDowell 1982).

asserted is that we can only grasp P-concepts if we *believe* that others instantiate P-properties.[4]

Given what I have said above, Stroud's objection seems exactly to the point. In order to justify the move from 8 to 9, it seems that Strawson must be relying on a conception of criteria that satisfies the two conditions mentioned earlier. But, not only is it difficult to see how criteria could satisfy both conditions, it would also seem that this conception of criteria as both required for understanding a concept and sufficient to secure knowledge of propositions involving that concept, is verificationist in some sense of that term.

In later work, Strawson (1985) accepts that Stroud's argument *does* in fact weaken his earlier case against other-minds scepticism. There he writes that,

> even if we have a tenderness for transcendental arguments, we shall be happy to accept the criticism of Stroud and others that either such arguments rely on an unacceptably simple verificationism or the most they can establish is a certain sort of interdependence of conceptual capacities and beliefs: e.g., as I put it earlier, that in order for the intelligible formulation of skeptical doubts to be possible or, more generally, in order for self-conscious thought and experience to be possible, we must take it, or *believe*, that we have knowledge of external physical objects or other minds. (Strawson 1985: 21–2)

Partly in response to Stroud's criticism, Strawson offers an alternative *naturalistic* answer to scepticism. The idea, which he finds in different ways in both Hume and Wittgenstein, is that certain beliefs are so fundamental to us that they are outside the realm of critical enquiry; they cannot be *reasonably* called into question. He suggests that the proposition that there are others much like ourselves may well be one of these. This is shown, as he sees it, by the fact that his (1959) argument against the other minds sceptic shows that we cannot but *believe* in the existence of others.

However, this view is problematic. For even if my belief that others exist is let off the hook of rational justification, it does not follow that my belief that *such and such a particular person instantiates such and such a particular P-property right now* is off that hook. That is, even if, as Strawson contends, we can legitimately ignore the question of how we can justify our general belief in others, it does not follow that we can legitimately ignore the question of how particular beliefs about others' P-properties are justified.

[4] 'for any candidate S, proposed as a member of the privileged class, the skeptic can always very plausibly insist that it is enough to make language possible if 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 1968: 255).

My belief that Paul is happy is not fundamental in the same way that my belief that other people exist perhaps is. And surely scepticism would be the result of a failure to offer an answer to any such particular question; the result would be that our ascriptions of P-predicates to others never do qualify as knowledge.[5]

There is, therefore, reason to see whether Strawson's argument can be modified in another way. What we need is a reading of Strawson's argument that does not rely on an intrinsically verificationist conception of criteria yet which has some real bite against the sceptic about other minds.

1.2 The modest reading

Strawson's primary concern is epistemological. That is, were the argument sound, it would provide us with an answer, for at least some types of P-property, to sceptical worries about other minds. But what is the scepticism that, according to the argument's conclusion, does not arise? It is useful here to distinguish between what, in another context, Martin (2006) refers to as 'Humean' and 'Cartesian' scepticism.[6] A sceptic of the Humean variety begins by pointing out that that of which we are aware in the perception of another falls short of that which we subsequently attribute to them. For instance, what I *observe* is various bits of behaviour (wincing, crying out, etc.), whilst what I *attribute* is pain. The Humean then goes on to argue that since there is no way for me to peek 'behind' this 'mere behaviour', there is no acceptable account of how I could *know*, rather than merely believe, that the other is in pain, or indeed is minded at all.

The sceptic of the Cartesian variety takes a different route to that same conclusion. He points out that for any of my veridical perceptual experiences of another as having some P-property or other, there could be a subjectively indistinguishable state which was nevertheless non-veridical. Thus, since I cannot ever tell which of these two situations I am in, I can

[5] Essentially this criticism can be found in (Sosa 1998). Strawson's (1998) response is unconvincing. It might be suggested that an answer to the latter question, analogous to Strawson's naturalistic answer to the former, may go something like this: we have an innate tendency to attribute certain P-properties on the basis of certain behaviour types, and that tendency cannot be rationally called into question. Perhaps there is some value in this suggestion, but it is pretty far from Strawson's claim and his original argument certainly does not support it.

[6] This should not be taken to suggest that any such positions were presented, let alone endorsed, by either Hume or Descartes. However, I hope that it will be obvious why these views are so named.

never know which I am in and so can never know whether another possesses some P-property or other, indeed I can never know whether that other is minded at all.

With this distinction in place, the reasoning of section 1.1 might be put by saying that the most that Strawson's argument could show is that scepticism of the Humean variety 'does not arise'. If, as Strawson argues, that which is observable constitutes 'logically adequate criteria' for the ascription of P-predicates, we can reject the Humean's argument at the very beginning. That of which we are aware in the perception of another does not fall short of that which is subsequently attributed. The simplest way for this to be true—the way that I earlier suggested that Strawson may actually have in mind—is to say that, in some cases at least, another's mentality is visible. That is, another's P-properties can enter into the content of one's perceptual states, and thus at least some P-predicates are non-inferentially ascribable.

But there appears to be nothing in Strawson's argument that would defeat or somehow undercut the Cartesian form of scepticism. On the face of it, Strawson's step 8 is entirely consistent with everything that the Cartesian sceptic asserts. That is not to say that the Cartesian line of reasoning should be accepted, just that whatever reason we have for rejecting it must come from elsewhere. Thus, although Strawson makes the perfectly general claim that scepticism about other minds 'does not arise', more realistically this should be limited to a certain type of scepticism, that which Martin calls Humean.

So, if we take Strawson's argument as employing a 'bland' conception of 'logically adequate criteria'—as signifying nothing more than a way of knowing that is not an inductive inference—and we rewrite the conclusion as,

10. Humean scepticism about other minds does not arise

then the argument may stand a chance of avoiding Stroud's charge of verificationism. Such a reading would nevertheless give the argument some real anti-sceptical bite, for a refutation of the Humean form of scepticism would be a genuine achievement. However, the conclusion would be modest at least in the respect that 10 entails neither that other minds exist nor that we know that they do.

There is even some, admittedly slender, reason to suppose that Strawson may have had this more modest reading in mind. First, as I have mentioned, it appears that Strawson would only be entitled to assume, as he

does in step 8, that signs and logically adequate criteria are exhaustive options, if he had the bland conception of criteria in mind. Second, also mentioned above, there is some evidence to suggest that Strawson accepts a perceptual account of our knowledge of other minds. If this constitutes at least one way that one can be in possession of criteria for the ascription of P-predicates, then there is some reason to believe that the criteria in question are not intrinsically verificationist. Third, after summarizing his argument, Strawson writes, 'The sceptical position customarily represents the crossing of the logical gap [between a mental state and its behavioural expression] as at best a shaky inference' (Strawson 1959: 109). Humean scepticism does indeed assume a picture according to which the move from behaviour to mental state is inferential. Cartesian scepticism, on the other hand, need not. Thus, there is some justification for thinking that Strawson has the Humean form of scepticism in mind. Having said this, whether Strawson intended his argument as ambitious or modest is not my main concern. For the remainder of the chapter I discuss the modest reading, referring to it simply as 'Strawson's argument'.[7]

Strawson's argument derives its anti-sceptical conclusion from a consideration of the necessary conditions of the possibility of self-consciousness (which I presume Strawson takes to involve the capacity for the self-attribution of P-properties). In particular, this is the role of step 5. I call any claim that purports to state some condition of the possibility of some form of cognition, a *transcendental claim*.[8] An anti-sceptical argument with a transcendental claim as a premise deserves the title of transcendental argument.[9]

[7] It should be obvious that the modest reading of Strawson's argument owes much to the reconstruction of Strawson in (Stern 2000: Ch. 6). I sympathize with much of Stern's discussion, in particular the argument against Ayer (1963). However, I depart from Stern in a number of ways. In particular, the distinction between Humean and Cartesian scepticism is not equivalent to Stern's discussion between epistemic and normative justificatory scepticism. Rather, both the Humean and Cartesian sceptic can be read as endorsing either epistemic or justificatory forms of scepticism. See (Stern 2000: Ch. 1.1). Further, in my §3 I find problems with even the modest reading of Strawson's argument that Stern does not.

[8] See the Introduction to this volume.

[9] Cassam (2007: 52–3) claims that since they present necessary conditions rather than 'means' of knowing, transcendental arguments cannot answer questions as to how knowledge is possible. One of the interesting features of Strawson's argument, as I have interpreted it, is that it shows that this distinction is not sharp. For Strawson's view is that a certain means of knowing (non-inferential) is a necessary condition of self-consciousness.

2. Objections to transcendental arguments

The reasonableness or otherwise of employing transcendental arguments has been much debated.[10] In large part, this debate focuses on transcendental claims. Transcendental claims, it is often supposed, are synthetic *a priori*, necessary, universal, and directed towards objective reality. Concerns can be raised with regard to each of these.

2.1 Synthetic a priority

Someone who offers an argument with a transcendental claim as a premise should be expected to argue for that premise. This, it might be thought, will be especially hard if those claims are both synthetic and *a priori*. Of course, within Kant's project, part of the purpose of the doctrine of transcendental idealism is precisely to allow for the possibility of synthetic *a priori* knowledge. Transcendental idealism provides one way of understanding how synthetic *a priori* knowledge could be possible. This fact suggests that a transcendental arguer who rejects transcendental idealism, a position that has tempted more than a few, notably Strawson (1966), must take one or other of the following options: treat transcendental claims as analytic; come up with some alternative account of synthetic *a priori* knowledge; or reject the terms in which I have set up this problematic, say by rejecting the analytic/synthetic distinction.

Whether this poses a threat to Strawson's argument depends on whether we should think of any of his premises as synthetic *a priori*. In fact, there is some reason to think that Strawson himself sees the all important step 5 as analytic. His case for the claim rests on a point that he calls 'purely logical': 'the idea of a predicate is correlative with that of a *range* of distinguishable individuals of which the predicate can be significantly, though not necessary truly, affirmed' (Strawson 1959: 99, n.1). If by 'purely logical' we understand 'analytic', the general concern over the legitimacy of synthetic *a priori* claims might be avoided. Of course, this is not to endorse either Strawson's purely logical point about predicates, or the suggestion that step 5 of his argument is a consequence of it.[11] But if those claims are suspect, it is not due to any purely general points about transcendental arguments.

[10] For very useful discussions see (Stern 2000) and the various essays in (Stern 1999).

[11] The latter is denied by Bermúdez (1998: 232–7).

2.2 Objective reality

As I have already mentioned, a well-known criticism of transcendental arguments is that associated with (Stroud 1968). Minimally expressed, this criticism can be viewed in the following way: if the form of a transcendental claim is 'Necessarily, If *A* then *B*', Stroud claims that *B* cannot plausibly be substituted by a claim about the way the world *is*, but rather must be limited to one concerning the way the world is *experienced* as being, or perhaps *believed* to be. Given the present interest in Strawson's argument, where the transcendental claim has the form 'Necessarily, everything is such that if it is *A* then it is *B*', the Stroudian claim would be that this should be replaced with 'Necessarily, everything is such that if it is *A* then it is experienced/believed to be *B*'. Only idealism, verificationism, or something equally poisonous could legitimate the stronger claim about the way the world *is*. Thus, if we wish to remain robustly realist, we must give up the pretensions of transcendental arguments to offer up conclusions directed towards objective reality rather than how reality is represented (either in experience or belief).

Stroud's argument is, I have argued, effective against the ambitious reading of Strawson's argument. But might it also have some force against even the modest version? Is there some claim that Strawson makes that is susceptible to a weakening from truth to belief? The obvious target for a Stroud-style objection would be step 5, yet such an objection looks to be inappropriate here. This is for the reason that Strawson is *not* saying that I must truly, or even actually, attribute P-properties to others. Rather, the necessary condition is that we have a *capacity* to make such attributions—we could make such attributions were we so minded and the conditions appropriate. It seems doubtful that Strawson's reasons for holding this true, rest on idealism, verificationism, or similar.

But perhaps this is a mistake. For a Stroud-style criticism of step 5 would maintain, not that our capacity for other-ascription may fail to yield true judgements, but rather that it might be only a *seeming* capacity, that we need have, in fact, no genuine such capacity, in order to be able to make self-ascriptions. This is fair, but implausible. The reason for this is that one's either experiencing or believing oneself to have the capacity to other-ascribe P-predicates is itself an exercise of that very capacity. Just entertaining the thought that one may or may not possess the capacity to ascribe P-properties to subjects other than oneself involves grasping the concept

of a subject *other than oneself*. If one has this concept, and some P-concept, then it seems difficult to see just how one could be denied the capacity to ascribe P-predicates to subjects other than oneself.[12]

Is this too quick? Have I not illegitimately moved from the capacity to *entertain the thought* that another possesses some P-property to the capacity to *judge* the same? For isn't Strawson's 'ascription' to be understood as a form of judging rather than mere thinking? This might seem especially worrying given that, on a Kantian view, there may be any number of thoughts—empty thoughts—that one can entertain, yet not be in a position to judge. I think, however, that this challenge can be met. I mentioned two possibilities above, belief and experience, each of which might be understood as involving the capacity to entertain the relevant thought. I think it should be clear that the concern does not arise in the case of belief. Believing that P is partly constituted by the disposition to judge that P, in the appropriate circumstances. Thus, if the Stroudian argument is that one need only believe oneself to possess the capacity to ascribe P-predicates to others, then this entails that one can judge that one can ascribe P-predicates to others. This, in turn, entails that one *can* ascribe P-predicates to others, although not necessarily correctly.

The case of experience is less clear. It is difficult to know what to make of a supposed *experience* of oneself as having the capacity to make other-ascriptions. We can, however, ask whether this experience entails the capacity to form experiential beliefs with matching content. Of course, this need not be the case if the experience in question is non-conceptual,

[12] It might be objected that I have wrongly identified exactly what capacity is at issue. Perhaps the capacity amounts to the ability to judge, based on what are taken to be logically adequate criteria. Then the Stroudian challenge is to show what, other than verificationism or some other view that ties the content of what we judge to the criteria by which we judge it, justifies our faith that what we take to be logically adequate criteria, really are. But, I don't think that this can be the right way of reading Strawson, since it entails that other-ascriptions based on less than logically adequate criteria are not exercises of the capacity in question. Strawson gives no indication that this is his view and, more importantly, his account of the relation between other-ascription and self-ascription appears to mitigate against it. Everything that Strawson says in this vein suggests that he supposes the ability to engage in the other-ascription of P-predicates, which can based on logically adequate criteria, to be part and parcel of the very same capacity as the self-ascription of P-predicates, which he claims are based on no criteria. He writes, 'In order to have this type of concept, one must be both a self-ascriber and an other-ascriber . . . In order to understand this type of concept, one must acknowledge that there is a kind of predicate which is unambiguously and adequately ascribable both on the basis of observation and not on this basis' (Strawson 1959: 108).

for then one need not possess the concepts required to form the belief. However, given that the supposed experience has such a sophisticated, high-level content (it is, after all, of oneself as having the capacity to make judgements about the mental states of others), I take it that a non-conceptual reading is implausible. Thus, I suggest that, if there is such a thing as having this experience, it involves the capacity to form the corresponding experiential belief and, therefore, judgement.

Thus, if we read Strawson's argument modestly, the Stroudian objection has no force against the transcendental claim made in step 5. This, I take it, is the most significant motivation for moving to the modest version of the argument in the first place.

2.3 Necessity

Transcendental claims are modal claims; they describe necessities. Different transcendental arguments support their transcendental claims in different ways, but there are at least three categories. First, a transcendental claim can be the consequence of a theory. Second, a transcendental claim might be based on a piece of conceptual analysis. Third, a transcendental claim might be (more or less explicitly) based on a claim about what is conceivable. Here it is argued that because it is not conceivable that P be true and Q be false, it is not possible that P be true and Q be false.

Transcendental arguments of the third variety clearly rely on some principle linking inconceivability and impossibility. But there is a question mark over the legitimacy of any such principle.[13] So before we accept that it is methodologically legitimate to base a transcendental claim on what is inconceivable, we must defend the principle that inconceivability is a good guide to impossibility.

This objection to transcendental arguments is distinct from the Stroudian objection.[14] Stroud's objection, if accepted, forces a change in what is claimed to be impossible. In the claim that *A* without *B* is impossible, *B* must concern the way that the world is represented as being. But this is still a modal claim. The present objection challenges even that weakened modal claim, contesting the transcendental arguer's right to claim any

[13] See the various essays in (Gendler and Hawthorne 2002). Also, cf. Dennett, 'Philosophers' Syndrome: mistaking a failure of imagination for an insight into necessity' (Dennett 1991: 401).

[14] Cf. the discussion in (Stern 2000: Ch. 2.3). Also see (Stern 2007).

modal knowledge whatsoever on the basis of the evidence of inconceivability. The problem raised by Stroud has become known as the 'inference to reality' problem. The present difficulty might be thought of as the 'inference to modal reality' problem.

If what I said earlier concerning the way in which Strawson argues for step 5 is correct, then it would seem that the claim is based on an analysis of the concept of a predicate. Since the claim in step 5 is based on conceptual analysis rather than an inconceivability claim, scepticism about the link between inconceivability and impossibility is neither here nor there. To this it might be responded that the activity of conceptual analysis involves reflection upon the nature of, and interrelations between, our concepts. Such reflection can tell us in which situations our concepts have application and in which situations they do not. In other words, conceptual analysis involves, at least in part, pushing our concepts to their limits in an attempt to determine whether a particular unusual situation really is conceivable. Thus, basing a transcendental claim on conceptual analysis and basing it on inconceivability are actually rather closely related.

If one were moved by such a thought, one might restate one's transcendental claim explicitly as a conceivability claim. That is, rather than claim that necessarily, if *A* then *B*, one could claim that it is not conceivable that *A* hold in the absence of *B*. Given this, the conclusion of the transcendental argument would have to be changed from *B* to, 'It is not conceivable that *B* is false.' Making this move has the virtue that it distances transcendental arguments from the controversy surrounding the relationship between inconceivability and impossibility. Indeed, if one's aim in presenting transcendental arguments is to elucidate the structure of our conceptual scheme, then such a limitation might not even be of special concern. For, one might argue, what is conceivable is given, at least partly, by our conceptual scheme itself. Thus, the way to lay bare the structure of that conceptual scheme is to give expression to those situations that are and those situations that are not conceivable. And, of course, this characterization seems to cohere well with the avowed intent of Strawson's descriptive metaphysics, within which his transcendental argument is situated. For, as Strawson famously wrote, 'Descriptive metaphysics is content to describe the actual structure of our thought about the world' (Strawson 1959: 9).[15]

[15] Also see (Strawson 1966: 271).

It may even be held that, limited to making claims about conceivability, transcendental arguments still have some anti-sceptical force. Suppose that *B* is the sceptic's target, that *A* is something that the sceptic cannot rationally doubt, and that *A* without *B* is inconceivable. It follows that, whilst the sceptic may be right that *B* is something we do not know to be true, its falsity is something that is inconceivable to us. This, it might be argued, is a serious blow to the sceptic's position.

In the case at hand, step 5 of the argument might be rewritten as,

5*: It is inconceivable that a subject possess the capacity to self-ascribe P-predicates yet fail to possess the capacity to other-ascribe P-predicates.

Of course, the argument's conclusion would have to be similarly adjusted, now maintaining that it is *inconceivable* that behaviour fail to constitute logically adequate criteria for the ascription of P-predicates, at least in some cases. As such, the reason that the sceptical question 'does not arise' would be that it asks us to contemplate an inconceivable state of affairs.

2.4 Universality

Transcendental claims are, at least in many cases, universal generalizations. Along with some of the issues discussed above, Sacks (2000) raises a worry about this feature. The concern is the relativistic one that transcendental claims, 'threaten to be only apparently universally necessary. The question is whether the modes of reasoning and the conclusions reached are merely historically or culturally indexed universalizations, such that however unimaginable it might be for those whose horizons are suitably set, there is nothing to preclude different universalizations equally well anchored elsewhere' (Sacks 2000: 285).

It is common to suppose that transcendental claims are universal quantifications. But there is some room for manoeuvre here. We need to specify the domain of quantification. At their strongest, transcendental claims would involve completely unrestricted quantification. So, Strawson's claim, 'necessarily, everything is such that if it can self-ascribe then it can other-ascribe', really would purport to hold of *absolutely every subject*. But one may baulk at making such a claim. Might there not be possible individuals so vastly different from us that their forms of self-ascription, of which we are entirely ignorant, fail to necessitate the capacity for other ascription? Perhaps. In such a case, one may wish to limit the domain of

quantification to subjects who are 'like us in some relevant respect'. For example, one might restrict the domain to 'discursive subjects'.[16] Given such a restriction, non-discursive subjects and their potential to surprise us with bizarre forms of consciousness, need not trouble our endeavour to discern the conditions of the possibility of (discursive) cognition.[17]

What would be the point of restricting the domain of quantification? Wouldn't it just be simpler to retain an unrestricted domain of quantification, but admit that the transcendental claim may not state a necessary truth? After all, it may still be true of all *actual* subjects. I take it that one of the main reasons for looking fondly on the strategy of restricting the domain would be the thought that one could thereby retain the status of the claim as *a priori*. For, although we no longer take this for granted, for the most part of the tradition of transcendental philosophy, to reject necessity is to reject *a priority*. If, then, we restrict the domain of quantification and retain the necessity operator, we can continue to maintain *a priority*. And retaining the *a priori* status of transcendental claims might be thought important insofar as our concern is to answer the sceptic.

Can Sacks' problem of universality be solved in this way? In Sacks' terminology, the question concerns the difference between genuine 'transcendental constraints' and mere 'transcendental features'. Introducing this distinction, he writes:

> Roughly, a transcendental constraint indicates a dependence of empirical possibilities on a non-empirical structure, say, the structure of anything that can count as a mind. Such constraints will determine non-empirical limits of possible forms of experience . . . A merely transcendental feature, on the other hand, is significantly weaker. Transcendental features indicate the limitations implicitly determined by a range of available practices . . . Human practices, concerns, interests, etc. will then stand to vary with empirical contingencies . . . Consequently, those transcendental features of what we can currently envisage are not constraints on what is possible. (Sacks 2000: 213)

[16] 'Kant's idealism depends crucially on his conception of human cognition as discursive . . . to claim that human cognition is discursive is to claim that it requires both concepts and sensible intuition' (Allison 2004: 12–13).

[17] The strategy of restricting the domain of quantification to subjects who are like us in some relevant respect should be treated with care. For what is then being claimed in the transcendental claim is that B is a condition of the possibility of *our particular kind* of cognition. One needs to take care so as not to reduce one's transcendental claim to what can look like the trivial consequence of an ad hoc move to avoid a counterexample.

A transcendental constraint is the consequent in, what I have been calling, a transcendental claim. A transcendental feature, on the other hand, is the consequent in a transcendental claim in which 'it is presently inconceivable that not' has replaced 'it is necessary that', *and* in which the universal quantifier has been restricted to that group of individuals that share our 'current practices'. Thus, the move from transcendental constraints to transcendental features involves rejecting both the necessity and unrestricted universality of transcendental claims. Given that Sacks goes on to call transcendental features, 'at best ordinary empirical constraints' (Sacks 2000: 214), I presume he means also to deny their status as *a priori*. If all we can lay claim to are transcendental features then, it might be thought, that is a serious blow to the pretensions of transcendental arguments. Restricting the domain of quantification to discursive subjects is one thing, restricting it to that group of individuals that share our 'current practices', is quite another.

At this point, we might remind ourselves that the target of Strawson's argument, on its modest reading, is Humean scepticism about other minds. However, Sacks' problem of universality has recourse to a far more radical form of scepticism. That is, it draws on the possibility of a general scepticism towards our 'modes of reasoning'. But Strawson might fairly adopt a principle of *one sceptic at a time*. Indeed, unless one thinks with, say, Husserl (1950) that the problem of others is conceptually or epistemologically *prior* to the problem of the 'external' world, then in even approaching the sceptical problem of other minds, one is already setting aside external world scepticism. Other people and their behaviour are, after all, part of the external world. I think, then, that it is legitimate to put aside scepticism about 'modes of reasoning', which in the current case is conceptual analysis, when considering Strawson's argument against the Humean other minds sceptic.[18]

3. Many questions

A more damaging criticism of Strawson's argument is that it muddles epistemic and developmental issues. In order to make this point clearly, I need to distinguish between a number of questions, any of which might

[18] I have not discussed Sacks' (2005) 'problem of transcendental necessity' which points out that Strawson's conclusion is that we have a certain capacity for identifying the P-properties of others if there are any others, but not that others must exist or even appear to. I take it that this problem is resolved once the move is made from the ambitious to the modest reading.

be being asked when we ask how knowledge of others is possible. First, there is the purely epistemic question,

K: In virtue of what do beliefs about another's P-properties, and the belief that there are others, count as knowledge?

Second, there is the question of methods,

M: What methods do subjects employ in the formation of beliefs about others' P-properties, and the belief that there are others?

Third, the developmental question,

D: How do subjects come to grasp P-concepts, and the concept of another subject?[19]

Strawson's argument has the overall form of a *reductio* of the initial assumption in step 1, that behaviour (that which is observable) is 'a sign' of the presence of P-properties (which are not observable). Let us call that assumption 'the sign view'. The sign view is most naturally understood as part of an answer to M, the question of methods. However, given that the notion of a sign is an epistemic one, it should also be read as part of an answer to the epistemic question, K. Now, the acceptability of step 2 of the argument implies that we are interested in only those versions of the sign view that rely on knowledge of correlations between behaviour and mentality. An actual example of someone who accepts something like this might be David Chalmers who claims that, 'We note regularities between experience and physical or functional states in our own case, postulate simple and homogenous underlying laws to explain them, and use those laws to infer the existence of consciousness in others. This may or may not be the reasoning we implicitly use in believing that others are conscious, but in any case it seems to provide a reasonable justification for our beliefs' (Chalmers 1996: 246). Chalmers explicitly distinguishes between K and M, offering something like a rational reconstruction of our knowledge of others. This suggests the possibility that Strawson's target is in fact the sort of account envisaged by Chalmers. Indeed, Strawson's use of 'know' in step 2 may suggest that he is interested in K, pursued independently of M. For, if K were not at issue, the occurrence of 'know' in step 2 would

[19] There is also the conceptual question, C: in what does a grasp of P-concepts, and the grasp of the concept of another subject, consist?

be unjustified, all that we would be justified in claiming is that one must *believe* or *assume* in some correlations between behaviour and P-properties.

But this cannot be correct. For, if the sign view were a view solely concerning rational reconstructions then there is no reason to suppose that it would conflict with step 5 of Strawson's argument, which places no requirements on justificatory relations. Strawson's claim is that the sign view violates the principle that self-ascribers must be other-ascribers. The sort of picture painted by Chalmers, however, can happily accept this, before going on to offer a rational reconstruction of our knowledge of others based on the observation of correlations between our own experiences and our own physical states. It seems, then, that Strawson *must* see the sign view as proposing an answer to M, and I think it is reasonable to suppose that his use of 'know' in step 2 is the result of an implicit presumption that questions M and K are to be pursued together.

But, now notice that step 3 of Strawson's argument assumes that we are interested in asking how subjects can have come to learn to ascribe P-predicates.[20] This is most naturally thought of as part of an answer to D, the developmental question. Strawson seems to assume that the sign view takes a stand on developmental issues. That this is so becomes evident when we reflect on the reliance on the notion of learning in steps 3 and 4. It is also clear from the fact that Strawson's argument (albeit tacitly) relies on step 7, the claim that the capacity for self-ascription could not be a condition of itself. On one reading, where 'condition' means *pre*-condition, 7 looks unarguable. Nothing can temporally precede itself. But on another reading, where 'condition' means necessary condition, 7 is evidently false, asserting that the capacity to self-ascribe couldn't be a necessary condition of the capacity to self-ascribe. To the contrary, everything is a necessary condition of itself. For the argument to be valid, 'condition'

[20] Actually, as I have quoted him, Strawson says 'know' not 'learn'. However, that he sees the sign view as involving an account of learning is clear from his ensuing discussion. There, in discussing the view that he has just rejected, he says that, 'there is not in general one primary process of learning, or teaching oneself, an inner private meaning for predicates of this class [a special subset of the P-predicates], then another process of learning to apply such predicates to others on the strength of a correlation, noted in one's own case, with certain forms of behaviour . . . [This picture is a] refusal to acknowledge the unique logical character of the predicates concerned' (Strawson 1959: 107–8). Also note the focus on the concept of learning in Stern's reconstruction of Strawson's argument (Stern 2000: 235–6).

needs to be read in the former, temporal, way.[21] So, once more, Strawson's assumption is that the sign view is committed to a particular account of conceptual development. I should stress that this is not to say that Strawson's primary concern is with developmental issues. It is not. Rather, my claim is that whilst Strawson's primary concern is with both the grounds for and the justification of ascriptions of P-predicates to others, he implicitly assumes certain developmental claims. In arguing that at least some ascriptions of P-predicates to others are grounded in logically adequate criteria and are thereby justified on non-inductive grounds, Strawson assumes that the contrary view is committed to a certain picture of the development of the capacity to make such ascriptions.

That picture, which I shall refer to as a 'first-person first' account, would involve the claim that subjects possess P-concepts and can apply them to themselves at a time prior to their being able to apply them to others. For it is only this sort developmental claim that would be inconsistent with the argument's all-important step 5, the transcendental claim linking the capacity for self-ascription with the capacity for other ascription. A first-person first account might say something such as the following: individuals acquire the concepts employed in attributing P-properties to others by noting, from their own case, correlations between behaviour types and mental P-property types, then observing behaviour elsewhere than their own bodies and making an analogy-based inference.

We should ask, then, whether on any natural reading of it, the sign view is committed to a first-person first account of conceptual development. I think it pretty clear that it is not. Consider, for example, the theory theory, which holds that attributions of P-properties to others are based on the application of a tacitly held folk psychological theory. Theory theory is a version of the sign view. Observed 'mere behaviour' is treated as one of the inputs to the tacitly held folk-psychological theory, with the attribution of P-properties being the output. Theory theorists may well think of the possession of a P-concept in terms of tacitly holding a theory that embeds that P-concept. Given that the folk psychological theory posited by theory theory is, indeed, a theory, the theory theorist may accept step 2

[21] Not quite. What is required is that 'is a condition of' be read non-reflexively. A non-temporal candidate might be the relation of rational justification. This would fit in with the reading of Strawson as attacking the sort of rational reconstruction view proposed by Chalmers. I have already said why I reject this reading.

of Strawson's argument, but will be quick to point out that such knowledge is tacit.

There are varieties of theory theory, two of which are the modular and the child-scientist versions. The modular view states that the tacitly held folk-psychological theory that is used in other-attribution is innate, whilst the child-scientist view holds that it is constructed by the child.[22] Supposing the theory theorist takes the innatist line,[23] he could reject Strawson's step 3, and so happily accept the all important step 5, whilst paying no attention to the rest of the argument. That is, he could endorse a version of the sign view that is not committed to a first-person first developmental account. Such an account would hold that the capacity to attribute P-properties to oneself co-matures with the capacity to ascribe P-properties to others. If this is so much as coherent, and I suggest that it is, then Strawson's argument fails.[24]

4. Naturalized transcendental arguments and scepticism about other minds

Strawson's argument assumes, incorrectly, that a proponent of the sign view *must* endorse a first-person first view of conceptual development. The mere fact that a version of the sign view that rejects a first-person first developmental account is so much as *coherent* is sufficient for Strawson's argument to fail. But what if it turned out that all such views were empirically false, that any empirically plausible version of the sign view was committed to a first-person first account? For example if, as Goldman (2006) argues on empirical grounds, the most plausible version of the sign view is a broadly simulationist one that is committed to the priority of first-personal ascription, might not something of the argument be salvaged? [25]

[22] See (Goldman 2006: Chs. 4 and 5) for useful, although critical, discussion.

[23] As is taken, for example, by Segal (1995). In actual fact, the important claim is not that the theory is innate, but that it is not acquired in the way that Strawson's argument assumes, via inductive generalization from one's own case. This wrinkle may turn out to be important depending on what we want to say about exactly how children's theories of mind change over time.

[24] Although I have presented this as an objection to the modest reading, it serves equally as an objection to the ambitious reading and also to the version presented in (Stern 2000: Ch. 6).

[25] Of course, Goldman would not put matters this way arguing, as he does, for the truth of his simulationist account.

The idea would be that since, empirically, the only plausible version of the sign view is committed to a first-person first developmental account, the sign view cannot tell the whole story. That is, since the capacity for self-ascription cannot exist prior to the capacity for other-ascription, behaviour must, at least in some cases, constitute logically adequate criteria for the ascription of P-predicates. Given its explicit reliance on an empirical premise, I think it reasonable to call this a 'naturalized' version of Strawson's argument.[26]

The addition of a non-self-evident, empirical premise to a transcendental argument, it might be argued, would mean that it could no longer provide any sort of answer to the sceptic. The necessity and, more importantly, *a priority* of transcendental claims is important, at least in part, to ensure that the premises in transcendental arguments are immune sceptical doubt.[27] An empirical transcendental claim, even if we did have evidence for its truth, would be wide open to sceptical doubt. Thus, the argument cannot be naturalized in this way. No amount of empirical evidence will help.

But this is too fast. Recall the *one sceptic at a time* principle. If one is engaging in a dispute with the other minds sceptic, *both parties* have, perhaps for the sake of argument, put aside scepticism about the 'external' world. So, by what right is the other minds sceptic simply able to help themselves to scepticism about the empirical evidence in question, say that of neuroscience and developmental psychology? Isn't this now moving the goalposts, introducing a far more wide reaching sceptical position? The defender of a naturalistic form of Strawson's argument can point out that the argument was never intended to counter such a form of scepticism.

At this point the sceptic might claim that the empirical evidence appealed to in ruling out all but first-person first versions of the sign view relies on the assumption that others exist, and it does so in at least two ways. First, the methodology of the cognitive sciences is not behaviouristic. Rather, cognitive science *assumes* that its subjects are experiencing subjects, not zombies. Second, our justification for believing anything that has been delivered via testimony relies on our justification for believing that those who testify are not zombies. Since the majority of the

[26] If Bell (1999) is correct, the sorts of transcendental argument pursued by Strawson are, in any case, embedded within a naturalistic framework that is quite un-Kantian.

[27] Cf. (Hookway 1999: 173).

information we acquire from the sciences is, in some way or other, testimonial, any reliance on scientific claims in an answer to other minds scepticism begs the question. If these two points are correct, then it seems that the sceptic is not changing the subject at all. Rather, they are simply drawing on the perhaps unappreciated consequences of their local scepticism.

Both points deserve more attention. The first makes a specific point about the human sciences. The second is far more general, ruling out reliance on any of the empirical sciences in an answer to scepticism about other minds. I begin with a few remarks about the plausibility of this second point. The question is whether I could know any testimony-based fact if I was not in a position to rule out the possibility of all humans other than myself being zombies. I think that there is a case to be made that we can. Zombie scientists are zombie-rigorous. Their experiments are zombie-controlled, and the quality of their work is subject to zombie-peer-review. Supposing that we are, or at least can be, justified in believing propositions based on the testimony of scientists, it is unclear why the fact that they might be zombie-scientists would make any difference. Further, the sceptic's position here would seem to make doubtful any knowledge gained from the use of computers which, just like zombies, we can assume to lack rationality and conscious awareness. This is highly implausible, especially given that the scientific evidence likely to be appealed to is such that it can, in principle, be checked oneself.[28] We do not, then, have a compelling reason to think that a defender of a naturalized version of Strawson's argument can appeal to no scientific evidence that he or she has not personally generated.

Of course, if the first point is correct, and one's justification for believing anything in the *human* sciences relies on one's justification for believing that *test subjects* are not zombies, then such acceptable evidence will be severely limited. To the extent that this point sounds plausible, and this is certainly the case with developmental psychology, the prospects for a naturalized version of Strawson's argument that does not beg the question against the other minds sceptic may seem dim. Dim, but not entirely dark.

[28] This may not always be the case. For example, there are certain mathematical proofs, such as the four colour theorem, that ineliminably involve computers. However, even in this case, the standard view is that we do know the theorem. Burge (1998) even takes such knowledge to be *a priori*.

In fact, much more needs to be said concerning the assumptions and methodology of each experiment and interpretation drawn upon in the case against co-maturational versions of the sign view. This, however, is not the place for such an examination.

5. Conclusion

I have argued that the modest reading of Strawson's argument can be defended against a number of general charges levelled against transcendental arguments. As it stands, however, the argument fails to answer the Humean sceptic about other minds. This is for the reason that it presumes the incoherence of versions of the sign view that reject the first-person first account of conceptual development. Perhaps this gap in Strawson's argument can be plugged empirically. However, given the fact that the cognitive sciences are not, in general, behaviouristic, it looks as though the argument may well prove resistant to such a naturalization.[29]

References

Allison, H. 2004. *Kant's Transcendental Idealism: An Interpretation and Defence,* revised and enlarged edition. New Haven: Yale University Press.

Ayer, A. J. 1963. The Concept of a Person. In his *The Concept of a Person and Other Essays*: 82–128. London: Macmillan.

Bell, D. 1999. Transcendental Arguments and Non-Naturalistic Anti-Realism. In Stern 1999: 189–210.

Bermúdez, J. L. 1998. *The Paradox of Self-Consciousness*. Cambridge, MA: MIT Press.

Burge, T. 1998. Computer Proof, Apriori Knowledge, and Other Minds. *Philosophical Perspectives* 12: 1–37.

Carruthers, P. and Smith, P. K. Eds. 1995. *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Cassam, Q. 2007. *The Possibility of Knowledge*. Oxford: Clarendon Press.

Chalmers, D. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. Oxford: Oxford University Press.

[29] Thanks to members of the philosophy departments at Essex and Manchester; to audiences in London and Cambridge; to two anonymous referees for OUP; and especially to Ann Whittle, Peter Sullivan, Lucy O'Brien, and Mark Sacks.

Dennett, D. 1991. *Consciousness Explained*. London: Penguin.
Gendler, T. S. and Hawthorne, J. Eds. 2002. *Conceivability and Possibility*. Oxford: Oxford University Press.
—— —— 2006. *Perceptual Experience*. Oxford: Clarendon Press.
Goldman, A. 2006. *Simulating Minds: The Philosophy, Psychology and Neuroscience of Mindreading*. Oxford: Oxford University Press.
Hacker, P. 1972. *Insight and Illusion*. Oxford: Clarendon Press.
Hahn, L. Ed. 1998. *The Philosophy of P. F. Strawson*. Illinois: Open Court Publishing.
Hookway, C. 1999. Modest Transcendental Arguments and Sceptical Doubts: A Reply to Stroud. In Stern 1999: 173–87.
Husserl, E. 1950. *Cartesian Meditations: An Introduction to Phenomenology*. Translated by D. Cairns. The Hague: Martinus Nijhoff, 1960.
Martin, M. G. F. 2006. On Being Alienated. In Gendler and Hawthorne 2006: 354–410.
McDowell, J. 1982. Criteria, Defeasibility and Truth. In his *Meaning, Knowledge, and Reality*: 369–94. Cambridge, MA: Harvard University Press, 1998.
Sacks, M. 2000. *Objectivity and Insight*. Oxford: Clarendon Press.
—— 2005. Sartre, Strawson and Others. *Inquiry* 48: 275–99.
Segal, G. 1995. The Modularity of Theory of Mind. In Carruthers and Smith 1995: 141–57.
Sosa, E. 1998. P. F. Strawson's Epistemological Naturalism. In Hahn 1999: 361–9.
Stern, R. Ed. 1999. *Transcendental Arguments: Problems and Prospects*. Oxford: Clarendon Press.
—— 2000. *Transcendental Arguments and Scepticism: Answering the Question of Justification*. Oxford: Clarendon Press.
—— 2007. Transcendental Arguments: A Plea for Modesty. *Grazer Philosophische Studien* 74: 143–161.
Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge.
—— 1966. *The Bounds of Sense: An Essay on Kant's* Critique of Pure Reason. London: Methuen.
—— 1985. *Skepticism and Naturalism: Some Varieties*. London: Methuen.
—— 1998. Reply to Ernest Sosa. In Hahn 1999: 370–2.
Stroud, B. 1968. Transcendental Arguments. *Journal of Philosophy* 65: 241–56.
Wright, C. 1980. Realism, Truth-Value Links, Other Minds and the Past. *Ratio* 22: 112–32.

Index