New Essays on the A Priori

Edited by

Paul Boghossian
and
Christopher Peacocke

CLARENDON PRESS · OXFORD
2000

reasoning. What sort of case have we been able to make for the claim that rule-circular arguments can provide genuine justifications for their conclusions? It seems to me that the case is substantial.

First, a rule-circular argument, unlike a grossly circular one, is not trivially guaranteed to succeed. Second, by relying on a small number of applications of a particular rule, a successful rule-circular argument delivers the conclusion that that rule is necessarily truth-preserving, truth-preserving in any possible application.[21] Finally, these applications of the rule will be applications to which the thinker is entitled, provided that the rule in question is meaning-constituting.

This case is constructed on the basis of several independently plausible elements. First, that the meanings of the logical constants are determined by their conceptual roles, and that not every conceptual role determines a possible meaning. Second, that if an inferential disposition is meaning-constituting, then it is a fortiori reasonable, justifiably used without supporting argument. Third, that something can be a warrant for something even if it is powerless to bring about a determined sceptic.

Putting these elements together allows us to say that we are justified in our fundamental logical beliefs in spite of the fact that we can produce only rule-circular arguments for them. The price is that we have to admit that we cannot use this form of justification to silence sceptical doubts. It is arguable, however, that, with respect to something as basic as logic, that was never in prospect anyway.

[21] Can we make sense of the idea that we are relying on only *one* application of a rule of inference? We do routinely discuss whether the application of a rule should be restricted—as when we debate whether MPP should be restricted in sorites cases—and that is enough to show that we *understand* what it would mean for a rule to apply in one context but not in another.

11

Explaining the A Priori: The Programme of Moderate Rationalism

Christopher Peacocke

I. INTRODUCTION

My starting point is a question about a distinction, a distinction between different ways of coming to know that something is the case. On a traditional rationalist conception, some ways of coming to know a proposition are justificationally independent of perceptual experience, while others are not. When you come to know a logical truth by way of your having a proof of it, you may need to perceive the inscription of the proof, and you may need various perceptual capacities to appreciate that it is a proof. But the justification for your belief in the logical truth is the proof itself. Perceptual experience gives access to the proof, which provides an experience-independent justification for accepting its conclusion. By contrast, if you come to believe 'That's Mikhail Gorbachev', when you see him at the airport, what entitles you to your belief is (in part) the perceptual experience by which you recognize Gorbachev. Your perceptual experience is not a mere means which gives you access to some experience-independent entitlement to believe 'That's Gorbachev.' This classical rationalist distinction between experience-dependent and experience-independent justifications or entitlements has been controverted, and objections to it raised and (in my own view) answered. Here I

Versions of this material were presented in 1999 to a symposium at the APA meeting in Berkeley, at which I was helped by the comments of co-symposiasts Laurence BonJour and Hilary Kornblith; to a seminar at New York University; to discussion groups in Oxford; and to the Princeton Philosophy Colloquium. My thanks to Tyler Burge and Michael Dummett for discussions of Fregean issues; and to Paul Boghossian, Bill Brewer, Justin Broackes, Bill Child, Michael Dummett, Dorothy Edgington, Kit Fine, Mark Johnston, Gideon Rosen, Ian Rumfitt, Stephen Schiffer, and David Wiggins for further comments and suggestions. In the early stages of writing, I had access to the superb research facilities of the Australian National University during a memorable visit there in January 1999, and was stimulated by discussions on these issues with David Chalmers, Frank Jackson, and Philip Pettit. Time was made available for work on this paper by a Research Professorship funded by the Leverhulme Trust; once again I thank the Trust for their invaluable support.

ask to be allowed, pro tem, to take for granted that the classical distinction is intelligible; and that the class of experience-dependent, and the class of experience-independent, entitlements are both non-empty. My opening question about the distinction is then as follows: what is the relation between those ways of coming to know a given proposition which are justificationally independent of experience, and the identity of the concepts in that proposition?

Propositions that can be known in a way which is justificationally independent of experience – propositions knowable in a way which is a priori, as I will say for brevity – seem to cover a vast range of subject matters. They include theorems of logic and arithmetic; they include the Gödel-sentence for any recursively axiomatized theory whose axioms are also known independently of experience; they include principles of colour-incompatibility; perhaps they include basic moral principles; perhaps they also include whatever truths of philosophy we know. So our initial question apparently concerns all the ways of coming to know that make available knowledge in these diverse areas.

The traditional rationalist answer to the question of what makes possible an a priori way of coming to know a proposition appeals to the notions of understanding and reason. Writers in the rationalist tradition, through Leibniz, Frege, and Gödel, have in one variant or another made such an appeal. They have held the view that it is in the nature of understanding certain expressions, or grasping certain concepts, that certain ways of coming to accept propositions containing those concepts are rational, and yield knowledge of those propositions, even when those ways involve no justificational reliance on perceptual experience. There are important and interesting differences between these rationalists; but that core seems to be common to them.

When one considers particular examples, the position common to the rationalists seems intuitive. When we assess the issues pre-theoretically, it does seem – for instance – that someone who possesses the logical concept of conjunction must be willing to make inferences in accordance with its introduction and elimination rules, without any need for experiential justification, and that these rules preserve not just truth, but also knowledge. Analogous points seem to hold for some basic arithmetical transitions. Similarly, it also seems that no more than grasp of the relevant colour concepts is required for one to be in a position to appreciate the incompatibility of a surface being wholly definitely red and wholly definitely green. Arguably similar points hold for the other examples too.

What has remained quite obscure in the rationalists' position, however, is the answer it would give to a how-question, the question 'How does understanding, or concept-possession, have this epistemological power? What is it about understanding which makes certain ways of coming to accept a given proposition yield knowledge, even though the way is justificationally independent of experience?' Many rationalists – including not only figures from several centuries ago, but also Gödel and Roger Penrose – have believed in a faculty of rational intuition or rational insight, a faculty which is supposed to explain the phenomenon of a priori knowledge. It is of course quite unclear how such a faculty is meant to work, how it could even provide truths about the world, let alone knowledge. The difficulties are difficulties of principle, and I will touch on them later. But the how-question which needs an answer is equally pressing for less exotic forms of rationalism. It is equally pressing for a less exotic writer such as Laurence BonJour, in whose writings the label 'rational intuition' is used merely to pick out the phenomenon of understanding-based a priori knowledge. BonJour's view is that it is 'anything but obvious' that the rational insight employed in attaining a priori knowledge involves 'a distinct psychological faculty'. BonJour holds that the psychological faculty involved in attaining a priori knowledge 'is simply the ability to understand and think'.[1] The more sober the view of what is involved in attaining a priori knowledge, the more challenging the task of explaining how understanding has the epistemological power. The more sober rationalist has to account for the epistemological phenomena from a non-exotic theory of understanding or concept-possession.

My goal in this paper is to make some proposals about the form such explanations should take, and to suggest some instances of the form in certain recalcitrant cases. A good understanding-based explanation of the capacity for knowing things by rational intuition should also have the resources for explaining why it is a fallible capacity, as it is widely acknowledged to be even by its most enthusiastic proponents.[2] If we can get a good account of the relations between understanding and a priori knowledge, we will be able to explain why some ways of coming to know a given proposition are a priori ways. If an a priori proposition is one which can be known a priori, this approach can also help to explain why any given a priori propositions has that epistemic status.

It is no part of this approach that meaning and understanding are involved only in outright a priori entitlements. On the contrary, it seems obvious that the identity of an observational concept is relevant to the issue of why it is that a thinker's perceptual experience entitles her to make a perceptual, empirical judgement. The task is rather to say what it is about understanding that makes a priori knowledge possible; which evidently does not preclude understanding from having a role in other ways of attaining knowledge too. It is also arguable that even in these empirical cases in which understanding is relevant to the status of something as a way of coming to know, a relative notion of the a priori has application. It is a priori, given the supposition that the subject is perceiving properly, and given the occurrence of a certain kind of perceptual experience, that a corresponding observational judgement will in those circumstances be correct.

The phrase 'rational intuition' has historically been associated with some of

[1] *In Defence of Pure Reason* (Cambridge: Cambridge University Press, 1998): 109.
[2] Ibid., sect. 4.5, 'The Corrigibility of Rational Insight'.

the headier forms of rationalism; but the phrase itself serves as a reminder for the sober too. The occurrence of the word 'rational' in the phrase emphasizes that the process of acceptance of a priori principles is a rational one. The occurrence of 'intuition' emphasizes that in many cases, the process of rational acceptance is not, or is not exclusively, one of derivation from axioms or principles already accepted. This apparently real combination – of rational acceptance which cannot be fully characterized as derivation from axioms or rules – is theoretically challenging for any rationalist, however modest and unexotic her theory of rational intuition. What we might call the *phenomena* of rational intuition which have been cited by the most interesting rationalist writers are extensive and striking. Whatever one thinks of Gödel's quasi-perceptual treatment of knowledge of the properties of concepts, the phenomena he cites – of rational acceptance of new axioms which do not follow from those previously accepted, of the notion of proof thus not being purely syntactically characterized, to mention just two – these are genuine phenomena which any good theory of understanding and the a priori ought to explain.[3]

II. MODERATE EXPLANATORY RATIONALISM

Now we can return to our opening question about the relation between ways of coming to know that are a priori, and the identity of the concepts in the content that is known. I distinguish two radically different general types of answer to the question that we can label respectively *minimalism* about the a priori and *moderate rationalism* about the a priori.

To formulate minimalism in this area properly, we need a distinction between composite and atomic ways of coming to know. One way of coming to know a logical truth is by working out a proof of it. The proof consists of a series of transitions, each one of which involves a way of coming to know a certain kind of conclusion from a certain kind of premises. The individual transitions at each line of the fully analysed proof involve an atomic way of coming to know, something that cannot be broken down further into other ways of coming to know. When you visually identify someone as a person who attended a course you gave some years ago, that can be broken down into constituent ways of coming to know. One constituent is your taking your perceptual experience at face value; another may be, for instance, your taking some memory image as of a student in your class at face value; and a third is your transition from the appearance of the face of the

[3] See esp. 'Russell's mathematical logic' and 'What is Cantor's continuum problem?', both repr. in *Philosophy of Mathematics: Selected Readings*, ed. P. Benaceraff and H. Putnam (Cambridge: Cambridge University Press, 1983, 2nd edn); and the philosophical papers in *Kurt Gödel Collected Works, III: Unpublished Essays and Lectures*, ed. S. Feferman *et al.* (Oxford: Oxford University Press, 1995).

person currently perceived and that of the remembered student to the conclusion that this is one and the same person. I will leave the distinction between atomic and composite methods at this relatively intuitive level for present purposes. It does need more elaboration, but this will be enough for a formulation of the core of the minimalist's position.

Minimalism is then the thesis that when an atomic way W is an a priori way of coming to know that p, it is simply primitively constitutive of the identity of one or more concepts in p that W is an a priori way of coming to know that p, or of coming to know contents of some kind under which p falls. That p can be known a priori in way W is, according to the minimalist, written into an account of understanding in the way that it is written into being a bachelor that bachelors are men, or, perhaps, written into being a chair that chairs have backs, or written into the relation of perception that a perceived object must causally affect the perceiver. It may be unobvious, and hard to discover, what is primitively constitutive of the identity of any given concept; but when one realizes that some property is so constitutive, there is no further answer to the question 'Why does that concept have that epistemic property?' According to this minimalist position, the fact that an atomic way W is an a priori way of coming to know that p is not consequential upon anything else. The minimalist will agree that those composite ways of coming to know which are a priori ways can be explained as such by being built up from atomic ways which are ways of coming to know a priori; but for the status of atomic ways as a priori ways, there is no further explanation to be given, beyond its being primitively constitutive of the identity of the concepts in the content known that they are so a priori. Perhaps the position would be better called 'epistemic conceptual minimalism', since the position employs talk of concepts without saying that such a talk is a mere manner of speaking. There are more radical forms of minimalism. But this epistemic conceptual minimalism seems to be minimalist within the class of positions which take at face value the talk of concepts and meaning in the theory of thought and understanding. What is important about the position is not that it regards the identity of concepts as in some cases given by the conditions for knowing certain contents, but that it regards the resource of what is primitively written into the identity of a concept as already a full explanation of the relation between the a priori and the concepts featuring in the content of a priori knowledge.[4] Such a minimalism remains a rationalist position, because it entails that the status of a way of coming to know as an a priori way traces back to what is, on the minimalists' view, involved in understanding and concept-possession.

There is really a cluster of positions which can be called (conceptually) minimalist. One variant of minimalism holds that when a thinker comes to know via

[4] That some concepts can be individuated by the conditions for knowing certain contents containing them was after all a claim of my book *Being Known* (Oxford: Oxford University Press, 1999).

an atomic a priori way that *p*, the thinker judges that *p* because of his grasp of the concepts in *p*. This statement is, according to this variant of minimalism, a genuinely explanatory true statement: understanding or grasp of the concepts in question explains the rational, a priori judgement that *p*. But that, according to this variant minimalist, is all that there is to be said on the matter.

The moderate rationalist, in the sense in which I will use that description, disagrees with all forms of minimalism. The first component of the moderate rationalist's view is that for any a priori way of coming to know a given content, there is a substantive explanation of why it is a way of coming to know that has a priori status, an explanation which involves the nature of the concepts in the given content. The moderate rationalist intends this claim to apply both to atomic and to composite ways of coming to know. For those who hold that concepts are individuated by the conditions for possessing them, this first component of the moderate rationalist's claim unfolds into the thesis that for any a priori way of coming to know a given content, there is an explanation of why it is an a priori way which has to do with the possession conditions of the concepts in that content.

The moderate rationalist is, then, committed to the feasibility of a certain explanatory programme. The goal of her programme is to identify those features of concepts which explain why a given way of coming to know a particular content is an a priori way. If the moderate rationalist thinks that concepts are individuated in terms of the conditions for their possession, execution of that programme must involve appeals to explanatory properties of concept-possession or understanding.

I am a moderate rationalist. Farther on, I will be suggesting ways in which we might make progress in carrying through the moderate rationalist's programme. But why should we prefer moderate rationalism to either variety of minimalism? I offer two arguments.

1. We already have some theoretical conception of understanding and meaning. For some of us, understanding consists in some form of grasp of truth-conditions; for others, a notion of canonical conceptual role is said to be basic. These theoretical conceptions of understanding and meaning are never put forward as merely partial determinations of understanding. These conceptions are not ones under which meaning can be fully characterized only by doing something else as well: by specifying additionally, as a further task, which ways of coming to know certain contents involving the meaning count as a priori ways of coming to know it. Once we have a conception of how meaning or content is determined, any links it has with the a priori have to be founded in that conception of how meaning or content is determined. If the links of meaning with the a priori cannot be so founded, one would not have fully explained meaning in terms of truth-conditions, or conceptual role, or whatever is the favoured notion. If meaning is already fixed as truth-conditions, or as canonical conceptual role, or whatever else is favoured, simply to add links with the a priori as further primitive axioms for the notion of meaning is simply to concede that meaning is not fully characterized without those extra axioms.

It may help here to consider a parallel with Dummett on the justification of deduction. Dummett insisted, rightly in my view, that deductive relations must be philosophically explicable in terms of the meaning of the logical constants involved in those relations.[5] A theory of meaning must explain why those deductive relations hold. This is a point which can be accepted by realists and anti-realists alike. The same applies to a priori ways in general, of which the deductive relations are a special case. If some principle has an a priori status, its status as such must be explicable in terms of the meaning of the expressions occurring in that principle.

There is an internal instability, a kind of unavoidable illusion, on the minimalist views. When we accept an a priori principle, it seems to be rational to do so, on the basis of our understanding of the expressions, or our grasp of the concepts, it contains. But this impression of rationality must be an illusion of one sort of another, on the minimalist view. If it were not an illusion, there would be some feature of meaning and understanding which explains a priori status. But that is precisely what the minimalist rejects. The variant minimalist who insists that grasp of concepts explains acceptance of a priori principles is not really in any better position. He has not accounted for the rationality of accepting an a priori principle.

2. This first point about meaning and understanding applies equally to the general concept of knowledge too. We have some theoretical conception of what is involved in a way of coming to accept a content being a way of coming to know it. The conception need not be reductive, of course. If there are principles connecting understanding with those ways of coming to know which are a priori ways, the connecting principles cannot have the status of primitive stipulations or axioms of the sort envisaged by any kind of minimalism. The connecting principles must have their source in the nature of knowledge, as well as in the nature of understanding, and the consequent relations between the two.

Such are the two initial, presumptive reasons in favour of developing a position which endorses the first component of a moderate rationalist's treatment of the a priori, the component which has a commitment to the possibility of explaining each case of a priori status by reference to features of understanding or concept-possession. But this first component of moderate rationalism cannot exhaust the content of any rationalism which is entitled to the label 'moderate'. For all I have said so far, an explanation of a priori status might invoke a theory of understanding which is quite extreme. A position of the kind sometimes attributed to Gödel,

[5] 'The Justification of Deduction', in *Truth and Other Enigmas* (London: Duckworth, 1978).

on which there is some faculty of rational intuition, allegedly analogous to perception, which puts a thinker in contact with concepts or meanings, could equally well endorse the existence of explanations which refer to the nature of meaning and understanding. In short: if the theory of understanding which proposes an explanation of the a priori status of a proposition is not itself moderate, the resulting position can hardly be a form of moderate rationalism either. So a second, obligatory, component of any rationalism which calls itself moderate must be the claim that the theory of understanding mentioned in its first component is not one which postulates causal or explanatory relations between properties of things in a third realm of concepts or meanings, and says that those relations are involved in understanding.

There is another reason of principle for wanting to include this second component. Could we attempt to explain a priori knowledge that *p* by some kind of causal explanation of the belief that *p* by the holding of the fact that *p*? Much of what is known a priori, including the mathematical and the logical, is arguably not of the right kind to enter causal explanations of mental states. But even in cases in which it is, there are reasons of principle for thinking that no such approach can explain the phenomenon of a priori knowledge. A priori positions hold in the actual world, however the actual world may be. That is, they hold fixedly, in the terminology of Davies and Humberstone.[6] Saying that the truth that *p* explains one's belief that *p*, and perhaps by some special causal route, fails to imply a crucial feature of the a priori, which is that *p* will hold whichever world is the actual world.

This objection to using causal explanation by the fact that *p* in the explanation of a priori knowledge that *p* roughly parallels the objection to using causal explanation by the fact that *p* in the epistemology of metaphysical necessity. Only what is actually the case – or , slightly better, only propositions whose truth is settled by what holds in the actual world – can enter causal explanations. The fact that *p*'s holding causally explains certain other events can never be sufficient for it to be necessary that *p*, just as it can never be sufficient for *p* to hold in the actual world, however the actual world may be.

In both the a priori and the modal cases, there will of course be causal explanations of why what is believed is believed, and these explanations can be of epistemic significance. The present point is only that what makes a piece of knowledge a priori cannot be fully accounted for by causal relations to what is known; and similarly for modal knowledge. It is good to be free of any commitment to causal explanation on that specific point, for attempts to develop the epistemology of the a priori or the modal in causal terms can only encourage the view that defenders of the a priori and of necessity must be committed to unacceptably non-naturalistic conceptions. One motivation for that charge, at least, is removed if our epistemology of these two notions is not causal.

[6] M. Davies and L. Humberstone, 'Two Notions of Necessity', *Philosophical Studies* 38 (1980): 1–30.

As I said, the Gödelian phenomena are genuine and important. Gödel is sometimes regarded as having a quasi-causal view of our relations to concepts and meanings, and if he did, he will thereby be open to the objections we have been raising. Faculties conceived by analogy with perception, far from helping to explain the possibility of rational intuition and a priori knowledge, are actually incompatible with the a priori status of the beliefs they deliver. However, nothing I have said rules out the more modest idea that there is a way of coming to know propositions which is a priori, is based on the understanding, and goes far beyond the models of understanding Gödel rightly criticized. I will return to this.

Our two-component moderate rationalist is making a highly general claim about all a priori truths, whatever their subject matter. The tasks for the moderate rationalist can be divided into four broad categories:

(a) There is an *identification task*. Identifying the way in which something comes to be known is often a hard part of the moderate rationalist's task. It is often highly plausible that there is an a priori way of coming to know a given proposition, whilst it remains obscure what exactly the way is. This is true of principles of colour incompatibility, of fundamental moral principles (if indeed they are a priori), and of some of the Gödelian phenomena. Saying that these principles are known by rational intuition cannot, for the moderate rationalist, be the end of the explanation.

(b) Once an a priori way of coming to know is identified, the moderate rationalist has then to explain why it is a way of coming to know a priori the content in question, on the basis of the nature of the concepts featuring in the content, and on the basis of their possession conditions if she conceives of concepts as individuated by their possession conditions. This is the *explanatory task*. It exists equally for the relatively a priori, as when we classify an inferential principle as a priori.

(c) The explanatory account of the a priori has also to be applicable to any a priori relation which is less than conclusive. This may include inductive principles, and principles about confirmation and probability. This is the task of *extension to non-conclusive cases*.

(d) A particular subclass of non-conclusive a priori principles of which the moderate rationalist must give an account are those stating that a subject is entitled to rely on the representational content of certain kinds of informational states in coming to make judgements. These cases include a thinker's entitlement to rely on perceptual experience in making observational judgements; on experiential memory in making certain judgements about the past; and so forth. An account of ways of coming to know that are a priori will not carry much conviction unless it can be extended to these cases too. I call this the task of *extension to reliance on informational states*.

This is a set of huge tasks; probably someone could spend a lifetime on them. Let us narrow our scope just a little, and consider the nature of the explanatory task (b) facing the moderate rationalist. From now on, I will also be considering only a moderate rationalist who holds that concepts are individuated by the conditions for possessing them (again, without any reductive presuppositions).[7] The explanatory task facing this theorist in a particular instance might concern the a priori status of a way W of coming to know some particular content containing essentially the concepts *C* and *D*. That is to say, the other contained concepts, unlike *C* and *D*, could be replaced uniformly by corresponding variables, and the universal quantification of the result could also be known a priori in way W. (I take the case in which only two concepts occur essentially, but these remarks will apply to any other number.) Our moderate rationalist's task can then be thought of as one of solving for, or discovering, a relation meeting certain conditions. The relation he has to discover we can call the *key* relation for the way W and the given a priori principle. It is a relation which holds between the following terms:

(1) the respective possession conditions for the contained concepts *C* and *D*;
(2) the semantic values of *C* and *D*; and
(3) the way W.

The key relation is one which explains why, when a thinker comes to believe the content in way W, he can know it to be true in the actual world, justificationally independently of perceptual experience.

The key relation which the moderate rationalist aims to find in any given case is one which will unlock the explanation of the a priori status of the given content. It can do so only if it is a relation between all of the elements (1) through (3). A priority is a phenomenon at the level of sense, not reference, and so on this moderate rationalist's theory must be traceable to the nature of the concepts involved. Hence element (1) must be present in the key relation, if concepts are individuated in terms of their possession conditions. If they are not so individuated, then it will be the nature of the concepts involved which forms the first element in the key relation, whatever that nature is. For element (2), we argue thus. Since, at least in the conclusive cases with which we are presently concerned, what is a priori is true, and indeed true however the actual world is, the semantic value of the concepts *C* and *D*, which contribute to the determination of the truth-value of the a priori proposition, must also be part of the explanation. Element (3), the way the thinker comes to make the judgement, when it is known a priori, must also be present, since the status of a belief as knowledge depends on how it is reached.

Finding something which is plausibly the key relation for a given way and

[7] That is, I will not be assuming that the conditions for possession can be given in the A(*C*) form of *A Study of Concepts* (Cambridge, Mass.: MIT Press, 1992). See the discussion in the later sections of my paper 'Implicit Conceptions, Understanding and Rationality', in *Philosophical Issues* 9 (1998): 45-88, ed. E. Villaneuva (Ridgeview: Atascadero, Calif.).

given principle is less challenging in one special case. That is the special case in which it is written into, or is a consequence of, the possession conditions for one or more concepts in the given principle that to possess those concepts, the thinker must be willing to accept the principle, by reaching it in that way. This seems to be the case for acceptance of the a priori principle that from A&B, it can be inferred that A, where the way in question is accepting inferences one finds non-inferentially compelling. The key relation for a special case like this is one abstracted from the following condition: that what makes something the semantic value of conjunction (viz. a certain function from pairs of truth-values to truth-values) is that it makes truth-preserving those inferences, like the inference from A&B to B, which are mentioned in the possession condition for conjunction, and which are made in the specified way. The key relation between the possession condition for conjunction, its semantic value, and a certain way of inferring something from it is the relation stated to hold in that condition. This is a key relation that brings in all the elements (1)–(3) we just identified.

This key relation explains the a priori status of the transition from A&B to B without postulating primitive, unexplained relations between understanding and the a priori, and without postulating problematic faculties. When the semantic value of conjunction is determined in such a fashion, the transition is guaranteed to be truth-preserving. It will be truth-preserving in the actual world, however the actual world is; that is, it holds Fixedly in the sense of Davies and Humberstone. This approach also explains why the way of reaching B from A&B should yield not just acceptance, but rather knowledgeable acceptance, of the transition. This method of reaching B, by inferring it from A&B, is immediately settled as correct in the actual world, however the actual world may be, given the contribution to truth-conditions made by the concept of conjunction. This goes far beyond brute reliability. The a priori correctness of the method is immediately founded in the nature of the contribution made by conjunction to the truth-values of thoughts in which it features. If that is not sufficient for a knowledgeable transition, it is not clear that anything could be.

The ordinary, non-philosophical thinker does not of course need to know the philosophical theory of why the inference is a priori. The philosophical theory is intended to explain why the ordinary thinker is entitled to make the inference from A&B to B without any justificational reliance on perceptual experience. A philosophical theory of the a priori will at many points have to use the distinction between entitlement, and what explains or grounds the entitlement.[8] The distinction is needed even in these very simple cases in which the transitions are written into, or consequences of, the possession conditions for the concepts involved in the transition. We must, in brief, always distinguish between

[8] For more on the distinction between justification and entitlement, see T. Burge, 'Content Preservation', *Philosophical Review* 102 (1993): 457–88.

(1) knowing, in a way which is a priori, that *p*

and

(2) knowing that it is a priori that *p*.

The ordinary thinker can know, in a way which is a priori, that *p*, without knowing that it is a priori that *p*. The initial goal of the moderate rationalist's programme is an explanatory characterization of (1), rather than (2); though of course it can also be expected to have consequences for (2) if the programme is successfully executed.

I mention this treatment of the very special case of ways of coming to accept principles which are written into the possession conditions for concepts in the principles in the spirit of offering an existence proof for the key relations of the sort to which the moderate rationalist is committed. In the remainder of this paper, I suggest some ways in which the moderate rationalist might attempt to carry through her programme in some more challenging cases. The more challenging cases are those of a priori principles which neither are, nor follow from, those principles mentioned in the relevant possession conditions. As always, the task for the moderate rationalist is to identify the key relation.

III. CONCEPTS TIED TO THE INDIVIDUATION OF PROPERTIES: TWO CASES

I turn to some case studies. Each case involves an example of a key relation which can explain epistemic phenomena that have been described by some as involving the use of rational intuition. Each of these rather different examples also illustrates a more general phenomenon: that of a concept being tied to the individuation of the property or relation it picks out. Here properties and relations are understood as being at the level of reference, but as more finely individuated than extensions or course-of-values. The different cases illustrate the different ways in which a concept can be tied to the individuation of a property or a relation.

First case: colour concepts and their a priori relations

Consider the colours red, green, blue, and the rest that are picked out by our ordinary colour concepts. Here I mean the colours themselves, not concepts of them, and not expressions for them. A colour's phenomenal properties are constitutive of it in at least the following respect. Take any particular finely discriminated colour shade *s*. This can be a shade as finely discriminated as Goodman would discriminate qualia: shade *s* is identical with shade *r* only if any shade matching *s* matches *r*, and conversely.[9] Here, as in Goodman, the range of 'anything' must

[9] See Goodman's criterion of identity for qualia in *The Structure of Appearance*, 3rd edn. (Dordrecht: Reidel, 1977).

be either universals, or at least something going beyond the range of actual particulars. Fix also on a given colour—red, say. Then if *s* is a shade which is clearly within the colour red, it is essentially and constitutively true of the colour red that *s* is clearly within it. (If *s* is a borderline case, that it is so is also essentially and constitutively true of the colour red.) The colour red is individuated by which shades fall within it, which fall outside it, and arguably by its pattern of borderline cases in respect of shades.

Since these phenomenal properties of the colour red are constitutive of it, they hold in all possible circumstances. It is a constraint on the genuine possibility of a world, or a world-description, that it respect the constitutive properties of objects, including colours. Hence: whichever world is the actual world, these phenomenal properties will hold of the colour red. They hold both necessarily and Fixedly. So far, these points all concern the level of reference, the level at which colours and shades themselves are located.

Now let us move to the level of concepts, sense and thought. The possession conditions for the concept *red* of the colour red are tied to these very conditions which individuate the colour red. Suppose a shade *s* is clearly a shade of red. If a thinker possesses the concept *red*, is taking his visual experience at face value, and if the experience represents an object as having shade *s*, then the thinker must be willing to judge 'That's red' of the presented object. We can relativize this to a part or region of the object; the point will still go through under such relativization. The thought 'That shade *s* [given in perception] is red' is not informative to the thinker who fully possesses the concept *red*.

Similarly, if a shade is clearly not a shade of red, the thinker must in those given circumstances be willing to judge 'That's not red.'

Next take a given shade *s* which is a shade of red and is not a shade of green. By the same reasoning again, applied both to the colour concept *red* and to the colour concept *green*, the thinker will be willing to judge, when taking perceptual experience at face value, when something is perceived as being shade *s*, 'That's red and not green.' The conditions for possessing the concepts *red* and *green* require the thinker to be willing to make this judgement; and it will be true.

It will also be relatively a priori that something with *that* shade (perceptually given) is red and not green. What I mean here by the claim of relative a priority is that the thinker's entitlement to this belief does not rely on the content of her perceptual experiences, beyond that content needed for having the relevant concepts in the first place. There is a way of thinking of a particular shade which is made available only by perceiving that shade. Such experience is necessary to have any demonstrative thoughts about that shade, including for instance such thoughts as 'That shade is or is not displayed on my colour chart', which are equally properly classified as relatively a priori. Such relatively a priori judgements contrast with 'The book with that shade is closed', which is not relatively a priori. What matters is that no further feature of the experience, beyond experience of the shade itself, is needed for the thinker's entitlement to judge, knowledgeably, 'That shade is red.' That judgement will hold whichever world is the actual world. (It will also hold necessarily.)

Now we can go for something more general. A thinker can reflect on what she can correctly judge when presented with a given shade. She can appreciate that if it is correct to judge, on the basis of perception necessary for having the demonstrative concept, something of the form 'That shade is a shade of red', it will also be correct, on the same basis, the make the corresponding judgement 'That shade is not a shade of green.' Suppose, what is also plausible, that every case in which something is red, or is not red, or is green, or is not green, could either be known to be so on the basis of perception; or else is a case in which something is counted as having one of these colour properties because it has the same physical properties which underlie the perception of colour in the perceptible cases. If a thinker can know all this, she can come to know that no perceptible shade is both a shade of red and a shade of green. Since the basis of this reflection is the relation of shades to colours which are in fact constitutive of the colours thought about, the generalization holds whichever world is the actual world. It holds fixedly that any shade which is a shade of red is not a shade of green. No particular course of perceptual experience is required to attain this knowledge: it is a priori.

This description of how such knowledge is attainable is founded in the possession conditions for the concepts *red* and *green*. Consider a concept whose possession condition is not tied to rational responses to the shades which individuate the colour to which the concept refers. For such a concept, it would not be possible for a thinker to appreciate such incompatibilities on a similar basis to that which we just outlined. Even if red is in fact the colour of the Chinese national flag, no merely understanding-based reflection could yield knowledge of the proposition that if a shade is of the same colour as that of the Chinese national flag, then it is not a shade of green. Such knowledge would have to be founded on the a posteriori, and not purely understanding-based, information that the colour of the Chinese flag is red.

What is crucial to this argument is the close relation between the way the colour is individuated, and the condition for grasping the concept *red* which refers to that colour. The relations to shades which contribute to the individuation of the colour are precisely those to which one who grasps the colour concept must be sensitive when making perceptually based judgements involving the concept. In short: the colour concept is tied to the individuation of its reference. It is only because this is so that a priori reflection on what it would be correct to judge in various circumstances can yield knowledge of colour incompatibilities.[10]

The need to invoke the tie between the colour concepts and the individuation of their references also seems to me to be one lesson of reflection on the early Putnam's discussion of colour incompatibilities.[11] Putnam's argument merits a paper-length discussion of its own: but to illustrate the lesson I just mentioned, I fix on the stage of his argument at which he writes 'And if it is true that no matter which shade of red and which shade of green we choose, nothing is both that shade of red and that shade of green, then it is true that 'Nothing is both red and green' even if by 'red' we mean not specific shades but broad classes of such shades' (1956: 211).

I say this by itself is not enough to explain a priori status. If the broad colour red is in fact the colour allowed by the local school for its dress code, it will equally be true that: no matter which shade of the colour allowed by the local school for its dress code and which shade of green we choose, nothing is both that shade of the colour allowed by the local school and that shade of green. But 'Nothing is both the colour allowed by the local school for its dress code and green' is not a priori. Putnam's principle needs some strengthening, some modal element, to get the stronger conclusion we need.[12] Putnam's principle is Fixedly true, and the prefixing of a 'Fixedly' operator gives the stronger premises. But then we have to ask: why are the stronger premises true? The answer I would give is that the concept *red* itself, unlike the concept *colour allowed by the local school for its dress code*, is tied to the individuation of the colour red. More specifically, this tie can be split up into several sublinks: the tie of *that shade* to a particular shade and what individuates it; the individuation of the colour in terms of its relations to the shade it includes; and the relation of the canonical broad colour concepts, *red, green*, and the rest, to the colours so individuated. So I think a fuller elaboration of this part of the early Putnam's position would need to draw on the resources I have been offering, and crucially on the notion of a concept being tied to the individuation of what it picks out.[13]

The explanation I have offered for a priori knowledge of colour incompatibilities, in being founded in the understanding-conditions for colour vocabulary, is one small step towards carrying through the moderate rationalist's programme. It

[10] I emphasize that I haven't shown that a material object has, at each point on its surface, only one colour. That would require further argument. All I have argued for is the a priori status of the proposition 'Any shade which is definitely a shade of red is not definitely a shade of green.' This would not be contradicted by the possibility of reddish-green, asserted by C. Hardin in his *Color for Philosophers: Unweaving the Rainbow* (Indianapolis: Hackett, 1988): 121–7.

[11] 'Reds, Greens and Logical Analysis', *Philosophical Review* 65 (1956): 206-17, and 'Red and Green All Over Again: A Rejoinder to Arthur Pap', *Philosophical Review* 66 (1957): 100–3.

[12] There is no modal element in his formalization of his argument later in the same paper.

[13] I differ from Putnam on some other points, particularly over what counts as a rule of language. Putnam gives a postulate which he says 'formulates a feature of English usage pointed out in the informal discussion: Nothing can be classified as both a shade of red and a shade of green.' This seems to me a truth about the non-linguistic world, not one about English. Insofar as the world cannot be a certain way, that will have consequences for which English sentences cannot be true—but the source of such impossibilities seems to me to have nothing to do with language at all. In his rejoinder to Pap, Putnam says 'it seems plausible to take 'Red and Green are different colours'' as 'direct linguistic stipulation' (1957: 102). I would contest this too: what is stipulated is which colour is the reference of the respective words; and then, given these referential stipulations, it's obvious with only a little thought (but not as a matter of any linguistic stipulation) that they are distinct. This is also what one would expect if understanding of colour words involves grasp of some class of paradigms and a closeness relation.

is a small step even within the special domain of colour. The moderate rationalist will also have the ambition of explaining all the other apparently a priori principles about colour which so intrigued Wittgenstein at different stages of his life.[14]

While colour concepts have their own distinctive properties, they are far from unique in having the crucial property of being tied to the individuation of their references. This more general property can explain other examples of the a priori, in accordance with the moderate rationalist's programme, as in the next example.

Second case: arithmetical relations

Consider arithmetical relations such as 'n is the sum of m and k' and 'n is the product of m and k.' At the level of the arithmetical relation itself, what it is for a triple of natural numbers to stand in these relations is given by their standard recursive definitions. But to think of these relations in the ways just given, as the sum relation and as the product relation respectively, is to have a fundamental method of calculating sums and products for which it is immediately obvious that it respects these recursions. So, for instance: the fundamental procedure for finding the sum of 7 and 5 involves counting up 5 steps from 7; and it is immediately obvious that this procedure respects the principle that 7 plus the successor of a number n is identical with the successor of the sum of 7 with n, i.e. that it respects the recursion for addition. A person may sometimes just see, or realize without conscious reasoning, that one number is the sum of two others; but if his judgement is queried, he must fall back on methods of calculating the sum of which it is clear (without substantial arithmetical computation) that they respect the standard recursive definition of addition. These latter methods are the thinker's fundamental procedures.

Judgements about the sum of two numbers, made by counting correctly, and without other mistakes, in the way one does in ordinary arithmetical calculation, will be correct in the actual world. They will also be correct whatever the actual world is like, because they involve thinking of these relations in ways tied to their very individuation. So these ways of coming to know the sums of two numbers are a priori ways of coming to know. The position is in agreement with Kant that 7+5 = 12 is a priori (though the reasons for this classification may not be the same). The a priori knowability of arithmetical sums is founded in the nature of the possession conditions for the concepts they contain, for they are tied to the individuation of the very relations the concepts pick out. A similar argument can be given for the a priori knowability of arithmetical multiplications, in relation to methods of calculating them involving addition.

A parallel argument can also be given about the relation between the individuation of the natural numbers themselves, and canonical concepts of them, if we regard conditions for application as partially or wholly individuative of the natural numbers. Once again, we first consider the natural numbers themselves, rather than concepts thereof. The number 0 is individuated by the condition that for there to be 0 Fs is for there to be nothing which is F. The number 1 is individuated by the condition that for there to be 1 F is for there to be something that is F and nothing else which is F. For any natural number which is the successor s(n) of some natural number n, the number $s(n)$ is individuated by its being such that for any property F, for there to be $s(n)$ Fs is (as Frege would have said) for there to be an object u such that the number of Fs other than u is n. The individuation of any number n in terms of the condition for there to be n Fs holds in the actual world, however the actual world may be. In this case, it is also necessary. This is still all at the level of reference, individuation, and metaphysics.

Then at the level of thought, to have a canonical concept c of some natural number n is to have a fundamental procedure for determining whether there are c Fs for which it is immediately obvious that the procedure respects the condition for there to be n Fs, the condition that is individuative of the number n. Counting is such a procedure. So the transition from the premises that the distinct objects x, y, and z are F, and exhaust the Fs, to the conclusion that there are 3 Fs, if the conclusion is reached by counting applied to x, y, and z, is an a priori transition. The transition is guaranteed to be true in the actual world, whichever is the actual world, because it is underwritten by what is individuative of the number 3.

This treatment of the case of numerical quantifiers can be combined with that of 7+5 = 12. We can thereby argue that the a priori status of 'If there are 7 Fs and 5 Gs, and nothing which is both F and G, then there are 12 things which are either F or G' can also be traced back to the phenomenon of concepts being individuated by their relations to the objects, properties, and relations they pick out.

IV. OBSERVATIONS ON THE PHENOMENON OF CONCEPTS TIED TO THE INDIVIDUATION OF OBJECTS AND PROPERTIES

I have been taking it throughout this paper that the a priori status of some content is a phenomenon at the level of sense. This may make it seem as if the a priori can only have to do with how we think of objects and properties. But when we realize that sometimes senses or ways of thinking are individuated by their relations to the very nature of what they pick out, it becomes clear that a priori truth can both be a phenomenon at the level of sense, and also have something to do with the nature of the objects or properties thought about. There is no incompatibility between those two characteristics.

The characterization of what it is for a concept to be tied to the individuation of its reference may make it sound as if this approach to such cases is committed to taking the ontology at the level of reference as somehow explanatorily prior.

[14] *Philosophical Remarks*, ed. R. Rhees (Oxford: Blackwell, 1975), and *Remarks on Colour* ed. G. E. M. Anscombe, trans. L. McAlister and M. Schättle (Oxford: Blackwell, 1977).

But that is not so. All that is needed in these philosophical explanations of certain cases of the a priori is a *link* between the concept and the individuation of the reference. That link can still exist even for a theorist who regards, say, the ontology of natural numbers or other abstract objects as some kind of projection of certain kinds of discourse, or modes of thought. That is certainly not a view I would recommend; but such a theorist would still have access to the present treatment of some cases of the a priori.

Second, I will briefly consider the relation between a principle's having an priori status because its constituent concepts are tied to the individuation of the properties and relations it picks out, and one of Frege's characterizations of apriority. In a famous passage, Frege wrote:

> It then depends on finding the proof and following it back up to the fundamental truths. If on this path one comes across only general logical laws and definitions, one has an analytic truth... But if it is not possible to carry through the proof without using truths which are not of a general logical nature, but belong to a particular domain of knowledge, then the proposition is a synthetic one. For a truth to be a posteriori, it is required that its proof should not go through without appeal to facts; that is, without appeal to unprovable truths lacking generality, and which contain assertions about particular objects. If on the contrary it is possible to carry through the proof wholly from general laws, which are neither capable of proof nor in need of it, then the truth is a priori.[15]

Few would want to argue that principles of colour exclusion are analytic in Frege's sense. But are such principles a priori by the characterization suggested by this passage? Is it possible to carry through proofs of them wholly from general laws which are neither capable of proof, not in need of it?[16] As Tyler Burge remarked to me, we have to take note of Frege's differentiation between the 'general logical laws' mentioned in Frege's characterization of analyticity, and the 'general laws', not necessarily logical, mentioned in Frege's condition for

[15] *Foundations of Arithmetic*, sect. 3, my translation (with some improvements thanks to David Wiggins). The original reads: 'Es kommt nun darauf an, den Beweis zu finden und ihn bis auf die Urwahrheiten zurückzuverfolgen. Stösst man auf diesem Wege nur auf die allgemeinen logischen Gesetze und auf Definitionen, so hat man eine analytische Wahrheit. ... Wenn es aber nicht möglich ist, den Beweis zu führen, ohne Wahrheiten zu benutzen, welche nicht allgemein logischer Natur sind, sondern such auf besonderes Wissengebiet beziehen, so ist der Satz ein synthetischer. Damit eine Wahrheit aposteriori sei, wird verlangt, dass ihr Beweis nicht ohne Berufung auf Thatsachen auskomme; d. h. auf unbeweisbare Wahrheiten ohne Allgemeinheit, die Aussagen von bestimmten Gegenständen enthalten. Ist es dagegen möglich, den Beweis ganz aus allgemeinen Gesetzen zu führen, die selber eines Beweises weder fähig noch bedürftig sind, so ist die Wahrheit apriori.' It is true that in this passage Frege is talking only about what makes a truth of mathematics an a priori truth. But no different criterion is suggested for other kinds of a priori truth; and his sufficient condition for being a posteriori is not confined to purely mathematical subject matter.

[16] I write 'suggested by this passage', because read strictly, Frege is here offering only a sufficient condition for a truth to be a priori. He may well also have believed it to be necessary. This is a complex and philosophically interesting question in Frege scholarship which I will have to forgo here.

apriority. Our question is to be understood as concerning the latter general, and not necessarily logical, laws.

The argument I offered earlier to the conclusion that no shade is both a shade of red and a shade of green relied on two assertions which Frege would classify as 'lacking generality'. It relied on the possession condition for the concept *red* of the colour red. That possession condition is not, as far as I can tell, a consequence of completely general laws alone. The argument also relied on principles about what individuates the particular colour red. These too are specific to the colour red. There was a dependence on the possession condition for the concept *red* in explaining the rationality (and relatively a priori character) of the transition from the experience, of any given shade which is clearly a shade of red, to 'That's a shade of red'. There was dependence on the individuation of the colour red in explaining why the argument is sound however the actual world turns out to be. There seems to be no satisfactory way to elaborate the soundness and a priori availability of this argument without appealing to truths about the particular colour red and the particular concept *red* (and, of course, their interrelations, which was the point of the preceding section).

It is true that I have relied upon a general philosophical theory of the way in which a relation between a concept and the individuation of the property or object it picks out can yield knowledge which is a priori. That general theory is formulated in terms which Frege would likely count as 'general laws'. But that general theory entails only conditionals of the form: if the relation between a property or object and a concept thereof is of a certain specified kind, then there will correspondingly exist a priori truths of a certain kind. To obtain specific truths which have an a priori status from the general theory, we need information about specific concepts, properties, and objects which are of the specified kind.

It has always been part of the rationalist position that understanding is a crucial resource in explaining a priori knowledge. Moderate rationalism is no exception. On the position I have outlined in the cases of colour and natural numbers, it is specifically a concept – that is, what is possessed in having understanding – which is tied to the conditions which individuate the object, property, or relation it picks out. But reflection on the quoted passage from Frege suggests that he may have been operating with a conception which recognizes three categories, of which, he seems to have held, only the first two may contain a priori truths. (a) There are the domain-independent logical laws. Arithmetical laws will reduce to these, in the presence of suitable definitions, if logicism is correct. (b) There are general laws which are domain-specific, but which are neither capable of proof nor in need of it. Frege famously held that geometry is a priori.[17] If the condition in the displayed passage is intended as a necessary, as well as a sufficient, condition of being a priori, Frege is thereby committed to saying that the axioms of geometry fall in this second category (b) (and thereby of course

[17] *Grundlagen*, sects. 87, 89.

acquires many a problem). (c) There are truths which are both domain-specific and specific to certain entities within that domain. Again, if Frege is offering a necessary condition of being a priori, then such truths as are in this category (c) will not be a priori under his account.

It is this last point that will elicit dissent on the part of the moderate rationalist who recognizes the consequences of the linking of certain concepts to the individuation of the properties or objects they pick out. That phenomenon generates a priori truths specific to particular concepts concerned with elements of a specific domain. The phenomenon is incompatible with simultaneous acceptance of the categorization (a) through (c), and of restriction of the a priori to the first two subcategories.

A rationalist may very reasonably want to distinguish between wholly general domain-unspecific principles and principles specific to particular subject matters. As far as I can see, however, there is no reason of principle to think that a priori knowledge must ultimately be explicable solely in terms of such general laws. There are even some reasons for doubting the coherence of such a position. For the same means by which one explains the possibility of a priori knowledge where it is not reducible to general laws also applies to general logical principles. What individuates a particular logical concept, whether one takes it to be a set of inferential rules, or an underlying implicit conception which specifies a contribution to truth-conditions, is arguably equally tied to the individuation of (for instance) a particular truth-function. If that is right, the a priori principles concerning specific objects or concepts come under the same general explanatory umbrella as the logical ones.

I also very briefly note the pertinence of the idea of something's being tied to the individuation of a property to the Kantian conception of pure intuition as an a priori means of establishing geometrical propositions. We can use some of the apparatus of this paper to give a limited defence of Kant's conception. Suppose just for this paragraph that we do not count imagination as experience, so that acceptance of a proposition on the basis of the deliverances of pure intuition could in principle at least be an a priori way of coming to know it. Pure intuition can be conceived of as a faculty which supplies representations whose content depends only on the constitutive properties of geometric objects – properties, lines, angles, and the rest. So one way of defending a neo-Kantian position about knowledge of pure geometry would be to note that in making geometrical judgements on the basis of the deliverances of pure intuition, one is being sensitive only to the constitutive properties of geometric objects. Judgements made on the basis of a proper exercise of pure intuition are thus a priori ways of coming to know. The relation of the faculty of pure intuition to what individuates geometrical objects is an essential component of the explanation of why this is so. One could develop this position without any idealism, transcendental or otherwise, and without any commitment to the a priori applicability of Euclidean geometry. Nor is the position one which embraces what I earlier rejected, viz. causal explanations of a priori knowledge by the truths known. The reason why judgements of pure geometry based on pure intuition will hold whichever world is the actual world is not (of course) that there is causal access to the non-actual. It is rather that only the properties and relations constitutive of geometrical objects are employed by pure intuition in the first place, when that faculty is properly exercised.[18]

V. A THIRD CASE: RATIONAL INTUITION AND IMPLICIT CONCEPTIONS

Some rationalists, and most famously Gödel, have pointed to phenomena which, they have said, we need rational intuition to explain. A brief list of these Gödelian phenomena would include the following.

(A1) We can have an understanding of some notions which goes beyond the axioms (or instances of axioms) we can write down for them. This is evidenced by the fact that we can discover new axioms which do not follow from those we previously accepted.[19] A fortiori, then, we can have an understanding of some notions which goes beyond the principles we must find immediately, and non-inferentially, compelling in order to possess those notions. This consequence as formulated is of course in my terms, rather than being a report of one of Gödel's theses. It is, however, a consequence all the same.

(A2) We can attain, in ways not based on sense perception, a better explicit statement of the nature of our notions, and thereby reach new axioms, which do not follow from those we previously accepted, and which are a priori truths.[20]

(B) It is part of the task of mathematics and logic to discover such new axioms or principles. Given that task, the use of rational intuition in these disciplines is ineliminable.[21]

(C) The notion of proof cannot be formalistically characterized. The notion of 'that which provides conclusive evidence' for a proposition cannot, even within the domain of mathematics, be something purely formalistic.[22]

(D) There may be finite, non-mechanical procedures which make use of the

[18] I believe the position outlined here is in the spirit of the remarks about the relation between geometry and a priori intuition by B. Longuenesse, *Kant and the Capacity to Judge* (Princeton: Princeton University Press, 1998): 290–1.

[19] Some sample passages from Gödel, amongst many others: 'Some basic theorems on the foundations of mathematics and their implications' (1995) 305, 309; 'The modern development of the foundations of mathematics' (ibid.) 385. All page references to these papers come from *Kurt Gödel Collected Works, III: Unpublished Lectures and Essays*, ed. S. Feferman, J. Dawson jun., W. Goldfarb, C. Parsons, and R. Solovay (New York: Oxford University Press, 1995).

[20] Sample passage: 'The modern development' (ibid.) 383.

[21] Sample passage: 'Is mathematics syntax of language?' (ibid.) 346–7; also in *Collected Works, III*.

[22] Sample passage: 'Undecidable diophantine propositions' (ibid.) 164.

fact that 'we understand abstract terms more and more precisely as we go on using them.'[23]

(A)–(D) do not exhaust what Gödel said about rational intuition. (D) also raises the issue of whether some sound procedures lie beyond the realm of the mechanical, an issue for which a full treatment would need some other occasion. But the Gödelian phenomena (A)–(C) are already ones which a philosophical theory of understanding and the a priori must explain somehow or other, if it is not (implausibly) simply to deny their existence. Gödel employed his underdeveloped quasi-perceptual theory of rational intuition in attempting to account for these phenomena. Though we may reject Gödel's philosophical account of rational intuition, it is, as Charles Parsons says, a 'real problem . . . for a theory of reason to give a better account'.[24] For the moderate rationalist, the challenge is to explain these phenomena by reference to properties of the (non-exotic) understanding involved in possessing the concepts involved in the axioms, proofs, and procedures Gödel is discussing.

More specifically, the moderate rationalist's task is once again to find in these Gödelian instances what I earlier called the key relation between the following three items:

(1) the possession conditions for the concepts in the axioms and principles Gödel discusses;
(2) the semantic values of those concepts; and
(3) the a priori way in which these axioms and principles come to be known in the Gödelian cases.

I want to suggest that the key relation for the Gödelian phenomena involves implicit conceptions of the properties and relations mentioned in the a priori axioms and principles.[25]

An implicit conception is, amongst other things, a content-involving subpersonal state, involved in fundamental cases in the explanation of a thinker's application of a given concept or expression to something. The content of the implicit conception specifies the condition for something's falling under the concept, or for the expression to be true of an object. To possess the concept, or to understand the expression, is to have the right implicit conception for it. Since possessing the concept and understanding are notions at the personal level, an implicit conception also has a characterization at the personal level. I would maintain that the implicit conception underlying the concept 'natural number' has the content given in these three familiar clauses:

[23] K. Gödel, *Collected Works, II: Publications 1938–1974*, ed. S. Feferman *et al.* (New York: Oxford University Press, 1990), at 306.
[24] 'Platonism and Mathematical Intuition in Kurt Gödel's Thought', *Bulletin of Symbolic Logic* 1 (1995): 44–74, at 64.
[25] For more on implicit conceptions, see my 'Implicit Conceptions, Understanding and Rationality'.

(1) 0 is a natural number;
(2) the successor of a natural number is a natural number;
(3) only what can be determined to be a natural number on the basis of (1) and (2) is a natural number.

As I emphasized in earlier work, it may sometimes be hard to articulate the content of an implicit conception underlying one of one's own concepts.[26] To articulate its content may be a major achievement. The process of articulation involves reflection and unification of the cases in which one knows that the concept does apply, and of the cases in which one knows it does not. Once the content of the implicit conception is correctly articulated, a thinker may be in a position to learn new principles, involving the concept he was employing all along, and which had not previously occurred to him. In the example of the natural numbers, one such principle would be the axiom that every natural number has only finitely many predecessors.

If it is true that to possess the concept *natural number* is to possess an implicit conception with the content (1)–(2), then this is also another case of a concept tied to the individuation of the property it picks out. For it is highly plausible that the conditions (1)–(2) also specify what it is to be a natural number, specify what is constitutive of that property.

Now we can take the Gödelian phenomena, starting with the fact the understanding of some notions outruns the general principles the thinker has so far written down for it, or even the principles he must be able to articulate in order to be credited with the relevant concepts. This is the Gödelian point I summarized in (A1) and (A2). If understanding sometimes consists in having an underlying implicit conception, then it is predictable that understanding in such cases may outrun the abstract schemata for the concept that one may be able to state. An account of understanding which appeals to implicit conceptions will already be opposed to the view of mathematics as 'syntax of language' to which Gödel was so opposed, and it will cite some of the same phenomena in its grounds for opposition.[27] Already in the humble case of the concept *natural number*, we mentioned a new a priori principle. Acceptance of the principle that any natural number has only finitely many predecessors is not something primitively written into possession of the concept of a natural number, along the lines minimal theories would have to propose. The principle is rather something whose correctness can be worked out by an ordinary user of the concept, on the basis of an understanding which is characterized without reference to that principle. As Gödel would say, it is an 'unfolding' of the concept we already had prior to formulation of the principle.

This explanation in terms of implicit conceptions does not require any appeal

[26] 'Implicit Conceptions, Understanding and Rationality'.
[27] See esp. the two versions of 'Is Mathematics Syntax of Language?' in *Collected Works, III*.

to a quasi-perceptual faculty of rational intuition to account for the phenomena. It is right to reject theories of understanding which cannot accommodate the phenomena, wrong to suppose that it is only the more exotic theories of rational intuition which can explain them.

Perhaps the most salient Gödelian case, for which the capacity for rational intuition has also been frequently and famously invoked by Roger Penrose, is that of the Gödel sentence *g* for a given recursively axiomatized theory with sufficient expressive resources to frame that sentence.[28] Penrose's view is that we can explain our knowledge that the Gödel sentence *g* is true only by appeal to what he calls 'mathematical insight'; and this, he says, eludes formalistic characterization. In 1990 he wrote,

> by the very way that such a Gödel proposition is constructed, we can *see*, using our insight and understand[ing—CP] about what the symbols in the formal system are supposed to mean, that the Gödel proposition is actually *true*! This tells us that the very concepts of truth, meaning and mathematical insight cannot be encapsulated within any formalist scheme.[29]

In *Shadows of the Mind*, five years later, he says that Gödel's theorem tells us 'the insights that are available to human mathematicians – indeed to anyone who can think logically with understanding and imagination – lie beyond anything that can be formalized as a set of rules. Rules can be a partial substitute for understanding, but they can never replace it entirely.'[30]

The significance of Penrose's argument, and thereby what is required to address the argument, has in my judgement been missed in the extensive discussions his argument has generated. There is a curious parallel here with the published discussions of John Lucas's partially similar views. I agree with David Lewis's remark that many of Lucas's critics have missed something important in Lucas's argument.[31] In the critics' rush to block arguments for the views which Penrose and Lucas reach, the critics have missed, and failed to address, important insights about truth and understanding which are involved in their respective cases. These oversights of many of the critics have then led Penrose and Lucas to think that their case is stronger than it really is. But let us move right to the core issue in Penrose's argument.

Suppose we accept some particular theory T, with a recursive set of axioms, and which includes first-order arithmetic. If we accept the theory, rationality requires us to accept that its axioms are true and that its inference rules are truth-preserving. So rationality requires us to accept that the theory is consistent. So far, it seems to me, this argument should be uncontroversial. Gödel gave a method of

[28] For a recent statement, see R. Penrose, *Shadows of the Mind: A Search for the Missing Science of Consciousness* (London: Random House, 1995).

[29] R. Penrose, 'Précis', in *Behavioural and Brain Sciences* 13 (1990): 643–705, at 648.

[30] *Shadows of the Mind*: 72.

[31] See D. Lewis, *Papers in Philosophical Logic* (Cambridge: Cambridge University Press, 1998): 166.

constructing a Gödel-sentence *g* for the theory, of which we can prove that if the theory is consistent, then *g* is not provable. This too is uncontroversial: these are theorems. So if we accept the theory, we are committed to holding that no number is the Gödel-number of a proof of *g*. By Gödel's method of construction, the sentence 'Every number is not the Gödel-number of a proof of *g*' is the sentence *g* itself. So we have just given an informal argument that, if we accept the original theory T, then it is rational to accept its Gödel sentence *g*, even though that sentence it is unprovable in T. This should still all be uncontroversial.

The controversy enters—or ought to enter—over the following issue: what is the correct account of our grasp of the meaning of universal quantification over the natural numbers, when it is such that we can appreciate the soundness of this reasoning to the truth of *g*? This is the crucial question on which discussions of Penrose's argument ought to be focused. Any treatment of Penrose's views, however convincing on other issues, will not have engaged with his argument unless it addresses the question of the nature of this understanding of such universal quantifications. Several writers have objected that nothing in Penrose's argument rules out the existence of an algorithm which correctly describes our mathematical reasoning, but which we cannot know to be such a correct description.[32] The objection is surely good; but it does not answer Penrose's question about understanding. Unless a better account of understanding is forthcoming, Penrose will go on thinking that a faculty of rational intuition involving mysterious relations to a Platonic realm is required for understanding; in any case, he will not have been given a full answer.

It seems to me that what is right in Penrose's argument is that the meaning of one or more expressions in the Gödel sentence *g* goes beyond anything wholly determined by the axioms and inference-rules of the theory T. If the meanings of all the expressions in *g* were fixed only by those axioms and inference-rules, it would be completely unexplained why *g* is true (let alone how we can know that *g* is true), since those axioms and inference-rules do not determine the correctness of *g*.

The critical question is then: what is the correct account of meaning and understanding for the expressions in *g*? I will be taking it that the important expression in this sentence for present purposes is the universal quantifier. The other expressions in the Gödel sentence are all symbols for primitive recursive functions and relations, whose meaning is plausibly fully determined by their standard recursive characterizations. We can, however, explain all three of our understanding of universal quantification, the truth of the Gödel sentence *g*, and our epistemic access to its truth under this hypothesis: that to understand the universal quantifier is to have an implicit conception with the content

[32] For a vivid statement of the objection, see H. Putnam's review of *Shadows of the Mind*, under the title 'The Best of All Possible Minds?', *New York Times*, Book Review Section, 20 Nov. 1994.

(U) Any sentence of the form 'All Fs are φ is true' if and only if every object of which F is true has the property expressed by φ.

The rational thinker familiar with the proof of Gödel's theorem knows that if the original recursively axiomatized theory T is consistent, then each of '0 is not the Gödel number of a proof of *g*', '1 is not the Gödel-number of a proof of *g*', '2 is not the Gödel-number of a proof of *g*', . . . is true. A thinker whose understanding of universal quantification consists in possession of an implicit conception with the content (U) can correctly and knowledgeably move from these ω premises to the conclusion that 'Every natural number is not the Gödel-number of a proof of *g*' is also true. Here I am taking it that the quantification over properties is unrestricted, and includes the property of being true.[33]

That (U) is the content of the implicit conception underlying this understanding is something which shows up in the pattern of cases in which a universal quantification is evaluated as true, and the pattern of cases in which a universal quantification is evaluated as false. The thinker may not yet have made explicit the content (U), though of course he may do so. It is not always necessary, in reaching knowledge based on possession of concepts underlain by an implicit conception, that the thinker himself make explicit the content of that conception (one can know that some seen object is a chair without being able to define the concept *chair* explicitly). The case also, incidentally, involves a second implicit conception—that underlying the concept *natural number*, which we mentioned earlier.

This account of knowledge of the proposition expressed by the Gödel sentence is in line with the moderate rationalist's programme. The account appeals to a property of the understanding of universal quantification, the property it has of consisting in possession of an implicit conception with the content (U). So on the present treatment, we have another case in which the phenomena which have been cited in support of rational intuition are genuine; but they do not require any exotic form of intuition or insight for their explanation.

One could imagine Penrose objecting to this account. 'Why', he might say, 'cannot we add the axiom (U), which you certainly accept as true, to the theory T, together with appropriate axioms for truth? When we do that, though, there will be a new Gödel-sentence g(TU) for the expanded theory; and how are you to give an account of truth, meaning and knowledge for that sentence g(TU)?'

Here, though, we must move very carefully. If we have a theory which includes disquotational principles for a truth-predicate, and is also capable of referring to all of its own expressions, we will be able to formulate the Liar Paradox, and the theory will be inconsistent. I think that (U), with its unrestricted quantification over genuine properties and meaningful expressions, does capture the content of the implicit conception underlying our grasp of universal quantification. (It does so in a non-reductive fashion, of course.) But precisely because of the problem of inconsistency, it does not follow that (U) can be embedded in just any theory which also uses a truth-predicate, and includes disquotation for all the sentences formulable in the language of the theory. The inconsistency stems not from any error in (U), but from permitting ungrounded uses of the truth-predicate in combination with an unrestricted principle of disquotation. The various known ways around this obstacle, some of which impose hierarchies, do not show that there is anything wrong with (U). Nor do they show that our grasp of truth is not capturable in general principles. They do not, in my view, even show that truth is an indefinitely extensible concept or property. They show only that one needs to take care in formulating a theory which contains a truth-predicate, contains axioms governing the truth-predicate, and also contains apparatus for talking about all of its own sentences, including those containing the truth-predicate.

[33] I rely on the natural extension (to the case of properties) of R. Cartwright's important defence of the legitimacy of such quantifications for the case of objects. See his 'Speaking of Everything', *Nous* 28 (1994): 1–20.

There is an argument that any view which tries to explain meaning proof-theoretically will have difficulty in giving a satisfactory answer to the Penrose-like question about meaning and understanding. It seems to me that the symbols for universal quantification over the natural numbers have exactly the same meaning when they occur in sentences formulable in some particular theory T, and as they occur in the Gödel sentence for T. Since the Gödel-sentence is a true universal quantification over the natural numbers which is not establishable by the methods of T, the person who accepts a proof-theoretical view of meaning must explain how such universal quantifications when they occur in sentences formulable in the language of T have the same meaning as the quantifier when it occurs in the Gödel sentence. The person who tries to explain meaning in proof-theoretic, or more generally evidential, terms may fairly say that the meanings are rather similar, though distinct. It seems to me, however, that the meanings are exactly the same.

The same challenge arises not only for proof-theoretical or evidential views. It also arises for a theorist who combines the following two theses. (1) He tries to explain the notion of natural number not modally, but has in place of clause (c) above some condition to the effect that the principle of arithmetical induction applies to the natural numbers; but (2) he does not think that one can quantify over all properties, but only over a limited totality. The theorist who holds this combination will also be vulnerable to the objection that he is forced to acknowledge only similar meanings, when there is really identity of meaning. For when the truth-predicate is added to a range of properties to which arithmetical induction can be applied, this theorist will have to say that the resulting 'new' notion of natural number and universal quantification over them is one for which new sentences can be shown to be true which could not be so shown on the old conception. But it seems to me, once again, that we ought not to say that the notion 'natural number' is ambiguous, or requires further determination.

I should also emphasize that there will be no problem for yet another theorist who follows the preceding theorist in respect (1), but does not follow him on (2).

This latest theorist prefers to explain the notion of natural number with induction written into the third clause of its characterization, but also permits the use of unrestricted quantification over properties, a range which will include the property of being true. This theorist will not suffer from any problem of ambiguity. This point also shows, incidentally, that the challenge from ambiguity to proof-theoretic views of meaning still arises, and can be answered, even if we do not use modality in the third clause of the characterization of a natural number.

The moderate rationalist will also agree with Gödel about (B), that is, that it is part of the task of mathematics or logic to discover new axioms or principles which do not follow from those we have already articulated, but which are nonetheless correct, and a priori, for the concepts we possess. We can learn, for a given subject matter, new a priori axioms which do not follow from those we already accept, and if 'rational intuition' is used to pick out the means, whatever it is, by which we attain such knowledge, then rational intuition is ineliminable in mathematics. Even in the case of axioms or axiom-schemata we already accept, their acceptability depends upon their respecting some meanings not explained in terms of acceptance of those axioms. We have to reflect and work out their correctness for the notions they contain, even if we have understood their vocabulary for many years.[34]

Gödel's third point (C), that the notion of proof cannot be purely formalistically explicated, will also be supported by the moderate rationalist. A formalistic account will not necessarily capture all the meaning-supported transitions which are sustained by the content of an implicit conception underlying a concept. (Even when a formalistic account extensionally captures them, what *makes* something a proof is its honouring of the links supported by that meaning.) Consider the ω-rule, that if F(0), F(1), F(2), are each provable in a system, then 'For all natural numbers *n*, F(*n*)' is provable. The moderate rationalist, again without any exotic claims about intuition, will say the following. Given the implicit conception underlying universal quantification, and the implicit conception underlying the notion of a natural number, the ω-rule is correct, and this is so not because it is a reasonable 'extension' of a meaning, or a further stipulation or determination of meaning, but because it is validated by the content of those two implicit

[34] For a suggested explanation of this phenomenon, see my 'Implicit Conceptions, Understanding and Rationality'. Gödel makes remarks which are consonant with this position, though what he says is not decisive against proof-theoretic explications of meaning. He writes 'It certainly looks as if one must *first* understand the meaning of a proposition *before* he can understand a proof of it, so that the meaning of 'all' could not be defined in terms of the meaning of 'proof'' (*Collected Works III*: 313). This is not decisive, since the proof-theoretical view can distinguish between implicit and explicit knowledge of meaning-determining rules. The proof-theoretical view may say that the prior understanding to which Gödel alludes involves only implicit, perhaps practical, knowledge of proof-theoretical role. Gödel goes on to add something more problematic for a proof-theoretical view: 'one may conjecture the truth of a universal proposition . . . and at the same time conjecture that no general proof for this fact exists' (ibid.).

conceptions underlying the concepts of natural number and universal quantification.

Another type of case in which Gödel notoriously wanted to apply his own conception of rational intuition is that of the rational acceptance of new axioms in set theory. He seems to have been optimistic that new, rationally acceptable axioms would eventually be found to decide the Continuum Hypothesis. I suspect many set theorists and philosophers of mathematics would agree with Charles Parsons's remark that 'The spectre of the concept of an arbitrary infinite set being a 'vague notion' that needs to be 'determined in a definite way' by new axioms isn't easily banished'.[35] It should not, however, be any part of the view of the moderate rationalist who implements his programme in some range of cases by appeal to implicit conceptions that the implicit conception underlying some particular concept of an abstract science should always be such as to determine the truth-value of such matters as the Continuum Hypothesis. What matters to the moderate rationalist are rather two points: (i) that there is a distinction of principle between those cases in which the underlying implicit conceptions are determinate in a given respect and those in which they are not; and (ii) the cases in which there is determinacy in a given respect can be used to explain some of the cases in which rationalists have made appeals to rational intuition. The moderate rationalist is not committed by the very nature of his position to endorsing the more optimistic estimates of which hypotheses and proposed axioms might eventually be decided by the implicit conceptions underlying set theory.

This moderate rationalist defence of a role for rational intuition may also part company at another point with Gödel if he held that the use of rational intuition in mathematics provides a kind of evidence that is unique to mathematics.[36] The idea of an implicit conception underlying a concept is entirely general, and can in principle be found in almost any domain, well beyond those of logic and mathematics. Implicit conceptions may underlie some observational concepts, some psychological concepts, some moral concepts, some political concepts, to name but a few other subject matters. Consider the following principles: for the observational concept *runs*, that when a person is running there is a moment at which both of the runner's feet are off the ground; for the psychological concept *being ashamed*, that shame about action requires identification with the person or institution whose action is in question. These principles can be informative to someone who has the concept. Explicit knowledge of them is not written into their possession conditions. Rather, they articulate one component of the implicit conception underlying grasp of the relevant concepts. Grander, though no doubt more controversial, examples could be

[35] 'Platonism and Mathematical Intuition in Kurt Gödels' Thought', 64.

[36] That it does so is either asserted outright, or attributed to Gödel, in R. Tieszen's excellent Critical Notice of Gödel's Collected Papers in *Mind* 107 (1998): 21932, at 230 (first sentence). If an attribution to Gödel is intended, Tieszen does not give chapter and verse.

given. In appealing to implicit conceptions to explain these phenomena of rational intuition, the moderate rationalist is assimilating rational intuition in the Gödelian cases to something much more general than the logical and mathematical. If we can attain it, a uniform explanation for the more general phenomenon seems to me desirable.

VI. CONCLUDING REMARKS

The scope of explanations of a priori status which appeal to concepts tied to the individuation of their references is not restricted to the cases I have so far discussed. Consider the principle-based treatment of metaphysical necessity, for instance, which I offered in *Being Known*. On that treatment, understanding metaphysical necessity involves having tacit knowledge of the conditions under which a putatively possible world-description represents a genuine possibility. That is, possession of the concept of metaphysical necessity is tied to the conditions which make something a genuine possibility. We can use this to explain the a priori status of certain principles of modal logic. If the thinker comes to accept those principles by drawing on the information he tacitly knows, and that information states what it is to be a genuine possibility, those principles will be guaranteed to hold in the actual world, whichever is the actual world. The case of metaphysical necessity also, incidentally, further illustrates the point that there will be cases in which a full explanation of the a priori will require one to address fundamental metaphysical issues about a domain.

In closing, I want to emphasize that in trying to delineate the phenomenon of concepts tied to the individuation of what they pick out, I have been concerned with only one species of explanation of a priori knowledge. Other examples of a priori knowledge have other kinds of explanation. It is equally the task of the moderate rationalist to supply these other kinds of explanation too. Indeed, reflection on the treatment I offered of the arithmetical case strongly suggests that other kinds of explanation must also exist. For that treatment presupposed the a priori existence of arithmetical objects, properties, and relations. A different, or at least an extended, model must be appropriate for explaining a priori knowledge of existence in the first place.

A second, very different, kind of case which shows the need for other kinds of explanation is that of 'I am here'. This case is one in which, it certainly seems, all the work in explaining why it is a priori is done in saying how the referents of the indexicals are picked out. The explanation has almost nothing to do with the nature of what is picked out, in strong contrast to many of the cases on which I have been focusing in this paper. The content 'I am here' is of course not a problem for moderate rationalism, but it does show that the moderate rationalist must acknowledge many subvarieties of explanation within the overarching conception

of the relation between concepts and ways of coming to know. I conjecture that in attempting to execute the moderate rationalist's programme over the full range of examples of the a priori, we will learn more about the many different ways in which a thinker may be related to the subject matter of her thoughts.