The John Locke Lectures 2009

Being Realistic about Reasons

T. M. Scanlon

Lecture 4: Epistemological Worries

I have claimed that there are irreducibly normative truths about reasons, and that the essential normative element in claims about reasons is a relation R(p, c, a) that holds between a fact, an agent in certain circumstances, and an action or attitude. The idea that there are truths about when this relation holds, does not, I argued, have puzzling metaphysical implications. To claim that something is a reason is just to claim that it bears this relation to some agent and action: nothing more. No metaphysically weightier property is required in order for truths about reasons to have the significance we attach to them. Nor, I have argued, is it puzzling why such facts should supervene on facts about the natural world.

But even, or especially, if truths about reasons represent a *sui generis* class of facts, distinct from natural facts, it may seem puzzling how we could come to know such facts. John Mackie, for example, claims that if we were aware of these facts "it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else." (38) He continues

> When we ask the awkward question, how we can be aware of this authoritative prescriptivity, of the truth of these distinctively ethical premises or of the cogency of this distinctively ethical pattern of reasoning, none of our ordinary accounts of sensory perception or introspection or the framing and confirming of explanatory hypotheses or

> inference or logical construction or conceptual analysis, or any combination of these, will provide a satisfactory answer; ‘a special sort of intuition’ is a lame answer, but it is the one to which the clear headed objectivist is compelled to resort.[1]

Mackie is here talking about *objective ethical* truths, but as I said in Lecture 2 I believe he would, or at least should, say the same about normative truths in general. The problem he sees arises from giving the idea of normative truth an unnecessary metaphysical reading, which brings with it the idea that knowledge of normative truths would require a special faculty analogous to sensory perception. I believe that this metaphysical reading should be rejected, and that the epistemological problem that it seems to bring with it is illusory, although there are other problems about knowledge of normative truths, which should concern us. As background for discussing both of these claims, it will be helpful to consider the case of mathematical knowledge, specifically knowledge of truths about sets. I believe that both the similarities and the differences between knowledge of truths about sets and knowledge of normative truths are instructive.

What gives rise to epistemological questions about empirical knowledge, and makes a causal theory of knowledge seem like an appropriate answer to these questions, is the fact that it is part of the content of most empirical judgments that they are about objects that are distant from us in space. If information is to get from them to us, how is this to happen but by their having a causal impact on our sensory surfaces? Transfer of

[1] Mackie, p. 39.

information by a non-causal process—some form of "intuition"—would be a strange and implausible alternative.

But things are quite different in the case of mathematical knowledge. Nothing in the content of mathematical judgments suggests that they are about objects with any particular spatial location at all, hence in particular not one "outside of us." Mathematical reasoning is about certain kinds of abstract structures and the relations between them. The conclusions of such reasoning have implications for things existing at particular places and times, such as about the number of pencils that would remain on my desk if I were to start with five and remove three. But the mathematical truth from which this particular empirical claim follows is a necessary truth about numerical quantities in general, which need have no spatial location.

The fact that mathematical facts and mathematical objects have no spatial location may be taken as ground for thinking that there is, after all, a special problem about mathematical knowledge. For if these objects exist "outside of space and time" the problem of explaining how information could get from them to us seems even greater than in the case of empirical truths. No causal link can bridge the gap, so some mysterious form of intuition seems to be required. But here the spatial metaphor has simply gotten out of hand. The idea of a region of existence "outside of space and time," and hence more inaccessible to us, is one we need not accept. If we reject this metaphor, however, we are still left with the question of how we discover truths about such matters.

There is not a *greater* epistemological problem about mathematical judgments than about empirical ones, but a problem or problems of a different kind. The problem is not how we could “be in touch with” the abstract structures that mathematics is about, but how we can characterize these structures in a way that makes clear which principles and modes of reasoning about them are valid.

Here the case of set theory provides a helpful example. Some axioms of set theory are generally accepted, but the subject matter of set theory cannot simply be identified with (the logical implications of) a particular set of such axioms, not even those of Zerrmelo-Frankel set theory, which are the most widely accepted. We need some basis for thinking that those axioms are correct: some way of thinking about sets and what sets there are that provides a rationale for these axioms and a basis for assessing other possible axioms as well. What kind of thinking can this be?

There is the question of whether further axioms should be accepted, and this question seems to be whether these axioms are *true*. To think about that question we need some way of thinking about sets that is independent of these particular axioms. And even the most widely accepted axioms would seem *ad hoc* if they were not thought to describe some determinate universe of sets.

So what kind of thinking is this? What are we doing when we are "thinking about sets" or about the concept of a set, if this is not a kind of perception? In a few cases it seems to be a matter of seeing what is true by definition, or "included in the concept of a set." For example, a set is defined as a collection of objects, the identity of which is determined by its members. So the axiom of extensionality, which says that no two distinct sets can have exactly the same members, seems true by definition. But most axioms are not of this character, even ones that seem entirely unproblematic. For example, the pair set axiom says that if a and b are sets, then there is a set c whose members are just a and b. This is not true by definition, or a conceptual truth, but it seems obviously true, because the way in which c is defined in terms of, or constructed out of, a and b is so clear and apparently unproblematic.

Consider one more example, the so called axiom schema of replacement. An instance of this schema says that for any set z, if some open sentence M(x, y) in the language of set theory such that $\forall x(x \varepsilon z \supset \exists !yM(x,y))$, then there is a set w which contains just those things that bear M to some member of z. (It is the set obtained by "replacing" each element of z with the set assigned to it by the relation M.) This way of defining, or constructing, new sets from given sets is not as simple as in the case of the pair set axiom, and its consequences are less obvious. But Replacement is widely accepted both because of its innate plausibility and because it leads to very plausible theorems, without, as far as anyone can tell after decades of use, generating any implausible conclusions, let alone contradictions.[2]

I take these to be good examples of how we come to have knowledge of sets. They seem to me to serve as a useful corrective to the tendency to think that such knowledge must come from one of two sources. Either we get it just by thinking about the concept of set, or we have access to it via some form of intuition of the realm of sets. The former limits set theory to "analytic" truths; and the latter seems mysterious. This might be called "the analytic/synthetic dilemma." It seems to me untenable, not because there is no distinction between the two possibilities it describes, but because each of these is misdescribed in crucial and misleading ways.

Some of the modes of reasoning abut sets that I have described may involve a kind of "picturing" (as in the case of the pair set axiom) but this picturing is not plausibly

[2] See Boolos, "The Iterative Conception of Set", p. 500; Parsons, *Mathematical Thought and its Objects*. Pp. 134, 339.

understood as involving “intuitive contact” with the realm of sets. It is rather a way of representing to ourselves ideas that we already have. All of the conclusions I described depend, ultimately, on Reflective Equilibrium thinking: all we can say is that they seem extremely plausible, and that there seem to be no obvious problems with their implications or with the line of thought that leads to them. So, even in the case of set theory, our conclusions are secured not, as Socrates would have it, by chains of logic, but by bungee cords of plausibility, of varying degrees of firmness.

The judgments of plausibility on which such thinking relies are ones we can be mistaken about. Naïve Set theory, according to which every open sentence Fx determines a set consisting of those things *a* such that F*a*, is initially very plausible. But it leads to paradox. Even leaving aside this outright contradiction, however, Naïve Set Theory might have been rejected simply on the ground that it allowed for the possibility that a set could be a member of itself, thus conflicting with the extremely plausible idea that the members of a set are “prior to” the set itself.

One thing that is attractive about the Naïve conception is that it provides a general characterization of what sets there are, in contrast to the piecemeal claims offered by particular axioms, which may seem ad hoc in the absence of some overall account. This raises the question of whether there might be some other overall account of the universe of sets that would provide a rationale for particular axioms.

One well-known response is what is called the Iterative Conception of Set.[3] This is a general characterization of the subject matter of set theory according to which the

[3] On this conception and the basis it provides for axioms of set theory, see J. R. Shoenfield, “Axioms of Set Theory,” in Jon Barwise, ed., *Handbook of Mathematical Logic* (Amsterdam: North Holland, 1977), George Boolos, “The Iterative Conception of

sets with which that theory deals are just those that would be formed in the following process: Begin, at stage 0, with a finite list of specified elements (or with the empty set.) At stage $n+1$ form all sets of the basic elements and the sets that were created at previous stages. For each limit ordinal λ, at stage λ form the set of all sets formed at stage α for all $\alpha<\lambda$.

The Iterative Conception could thus be seen as naturally arising from this conflict in a process of seeking Reflective Equilibrium, that is to say, a process of finding general principles that are compatible with our considered judgments.This provides reason to look for a general characterization of the universe of sets that might provide a rationale for currently accepted axioms and perhaps a basis for arguing about additions to them. One attempt to do this is what is called the Iterative Conception of Set.[4] This is a general characterization of the subject matter of set theory according to which the sets with which that theory deals are just those that would be formed in the following process: Begin, at stage 0, with a finite list of specified elements (or with the empty set.) At stage $n+1$ form all sets of the basic elements and the sets that were created at previous stages. For each limit ordinal λ, at stage λ form the set of all sets formed at stage α for all $\alpha<\lambda$.

Set," *The Journal of Philosophy* 68 (1971) pp. 215-231, and "Iteration Again," *Philosophical Topics* 17 (1989) [both reprinted in Boolos, (R. C. Jeffrey, ed.) *Logic, Logic, and Logic* (Cambridge, MA: Harvard University Press, 1998).] For discussion see also Charles Parsons, "What is the Iterative Conception of Set?" in his *Mathematics in Philosophy* (Ithaca, NY: Cornell University Press, 1983).

[4] On this conception and the basis it provides for axioms of set theory, see J. R. Shoenfield, "Axioms of Set Theory," in Jon Barwise, ed., *Handbook of Mathematical Logic* (Amsterdam: North Holland, 1977), George Boolos, "The Iterative Conception of Set," *The Journal of Philosophy* 68 (1971) pp. 215-231, and "Iteration Again," *Philosophical Topics* 17 (1989) [both reprinted in Boolos, (R. C. Jeffrey, ed.) *Logic, Logic, and Logic* (Cambridge, MA: Harvard University Press, 1998).] For discussion see also Charles Parsons, "What is the Iterative Conception of Set?" in his *Mathematics in Philosophy* (Ithaca, NY: Cornell University Press, 1983).

The iterative conception is an attempt to spell out the idea of priority, omitted from the Naïve Conception, the idea that a set is an "arbitrary collection of *pre-existing* elements." It thus begins from a idea that might be thought of as part of the concept of set but goes beyond what could be claimed to be contained in that concept, to make further claims that are, however, extremely plausible, if not entirely unproblematic.[5]

The sequence of thinking that I have just described is naturally seen as a process of seeking Reflective Equilibrium. We begin with a claim, the Naïve Conception, that seems very plausible. Investigating its consequences we see that it leads to unacceptable results. Looking back, we can see that these flow from a mistake in our original thinking: there is a way in which the Naïve Conception should not have seemed so plausible, because it fails to do justice to the idea that the members of a set are "prior to" the set itself. Taking this into account, we formulate a new claim, the Iterative Conception, and investigate *its* implications.

The characterization of the realm of sets provided by the Iterative Conception is, however, incomplete. It provides a rationale for most of the axioms of Zermelo-Frankel set theory, but it does not provide support for the axiom schema of replacement. (Instances of that schema may be true of the universe of sets described by the Iterative Conception, but the truth of these schemata does not follow from that conception.) As I have said, Replacement is nonetheless widely accepted because of its plausibility, usefulness, and the plausibility of its consequences. (More reflective equilibrium thinking.)

[5] As Parsons notes, there are difficulties about how the idea of "earlier and later" in the sequence is to be understood. See his, "What Is the Iterative Conception of Set?"

Two further points about the Iterative Conception will be helpful in what follows. The first is that although the Iterative Conception is “external” to any particular axiomatized set theory, it is itself a piece of mathematical thinking, a description of an abstract mathematical structure. It employs the notion of set, it appeals to the idea of “*all sets* of elements formed at previous levels,’ and it relies on the idea of transfinite ordinals. Despite this, it is a useful characterization. As Boolos writes,

> It is not to be presumed that the concepts of *set* and *member of* can be explained or defined by means of notions that are simpler or conceptually more basic. However, as a theory about sets might itself provide the sort of elucidation about sets and membership that good definitions might be hoped to offer, there is no reason for such a theory to begin with, or even contain, a definition of ‘set’. That we are unable to give informative definitions of *not* or *for some* does not and should not prevent the development of quantificational logic, which provides us with significant information about these concepts.[6]

I believe, then, that there is no general epistemological problem about set theory, arising from the difficulty of explaining how we could get in touch with abstract objects. The epistemological questions that need answers are of two kinds. The first question is how to understand the forms of reasoning that can lead to particular valid conclusions about sets. The second question, related to but not identical with the first, is whether, or in what cases, we have reason to believe that all statements about sets have determinate truth values. (This is analogous to the question, discussed at the end of my last lecture, about the idea of correctness that is applicable to claims about reasons.)

[6] “The Iterative Conception of Set,” in Benacerraf and Putnam, p. 486.

An answer to the first question is implied by what I have already said: we can reach conclusions about sets by a combination of Reflective Equilibrium reasoning and deductive argument from premises supported by such reasoning. The Iterative Conception plays a role in this process by providing a partial characterization of the realm of sets that provides a rationale for some axioms, thereby, one hopes removing the suspicion of arbitrariness. The Iterative Conception may also be helpful in answering the second question, of determinateness, and this is the final point I want to make about it.

The description of the Iterative Conception relies at crucial stages on the idea of "all sets of items formed at previous stages," and this may be thought to leave open important questions. For example, does "all sets" include, for any previously formed set A of disjoint sets, a "choice set" C containing one member of each of these sets? The description of the Iterative Conception that I have given leaves this open, and any way of extending it to settle the matter seems to beg the question.[7]

This might be taken as further evidence that the Iterative Conception fails to characterize adequately our intuitive idea of "all sets." But does this intuitive idea itself identify a universe of sets in which there is a determinate answer to the question of whether there will always be such a choice set? Consider the lines of thinking leading to different answers as to whether such a set will always exist.

The line leading to a positive answer begins from the thought that the description of a choice set does not seem problematic. Why shouldn't "all sets" include, for each A, at least one such set? Moreover, it might be added, the thesis that there always is such a

[7] Boolos, the Iterative Conception of Set," pp. 501-502.

set, called the Axiom of Choice, is so useful in proving further theorems that it is widely used and has not led to contradiction.

The contrary view begins by noticing that the description given of the choice set C did not identify a particular set. This is in contrast to the other axioms we have considered, the axiom of Pairs and the axiom schema of Replacement, which identified, for any given sets a particular new set constructed from them. Things would be different if we could assume that all the sets in A were well ordered. Then the choice set C could be defined as containing the first element of each member of A. Without this, all we have is a description of a *kind* of set, and all we can say is "Why shouldn't there be a set of this kind?

A constructivist like Michael Dummett would say that in the absence of a proof that there is a particular set of this kind, or a proof that there is no such set, the question of whether such a set exists has no determinate answer. More generally, a characterization of the universe of sets such as that given by the Iterative Conception does not guarantee that every well formed statement in the language of set theory is either true of false of this universe (independent of whether we have any way of determining which.)[8] A moderate conception of what determinateness requires might hold that statements about sets have a determinate truth value only if *there is* some proof that settles this question (whether we now have that proof in hand or ever will discover it.) Dummett's constructivism is more demanding. He holds in order for a mathematical

[8] See, for example, his essay, "Platonism," in *Truth and Other Enigmas* (Cambridge, MA: Harvard University Press, 1978), pp. 202-214.

statement to have a determinate truth value we must have in hand some proof which establishes it.[9]

Dummett’s position is called anti-realism, which makes it sound like an ontological position. But the dispute here is not, I would say, an “external” ontological one about the existence of abstract objects. The issue is rather about the kind of reasoning that is required to support the claim that a mathematical *statement* has a determinate truth value. What it leads to is not a kind of fictionalism, or some other denial that statements in set theory, say, are ever true or false, but a particular view about what constitutes adequate mathematical reasoning.

Let me return now from these reflections on the philosophy of set theory to questions of practical reasoning, which are my concern in these lectures. There are important differences between set theory and practical reasoning. The subject matter of set theory is an abstract theoretical domain which we can characterize in a way that makes it possible to argue about it in a precise and formal manner. The subject matter of practical reasoning is, as the name implies, practical, and it is much less precise, perhaps incapable of being rendered so.

I do not mean to minimize these differences. I have discussed set theory in order to make several points which I believe are instructively analogous to the case of practical reasoning. The first of these is a point about method.

Practical reasoning is like set theory in not being about the physical world. But, just as in the case of set theory, insofar as there is an “epistemological” problem about our knowledge of truths about reasons it is not a problem generated by the fact that we

[9] See “Realism,” in *Truth and Other Enigmas*, p. 164.

are not in causal contact with the relevant facts. Our conclusions about reasons for action depend, ultimately, on Reflective Equilibrium thinking: all we can say is that they seem plausible on reflection, and that there seem to be no obvious problems with their implications or with the line of thought that leads to them. At the most abstract and general level, this is enough. The epistemological problem about practical reasoning is whether this reflective equilibrium process leads us to a sufficiently clear and determinate characterization of the kind of reasoning that supports these conclusions.This is the question of determinateness, which I have just been discussing. The problem is more difficult in the case of practical reasoning than in that of set theory because of the relative weakness of the relevant forms of reasoning.

One question for us, then, is whether we might find a general characterization of reasons for action that would have this role, in particular, a general characterization that would support the idea that practical reasons constitute a determinate domain about which there are definite answers. What I called in my first lecture a normative desire theory is a candidate for this role. As I pointed out then, this theory is itself a normative thesis, so it would not answer metaphysical worries about normative truths, or epistemological worries derived from them, nor does it explain the practical significance of judgments about reasons. I have argued, however that these problems are not real. The Iterative Conception of Set also provides a useful example here, of how a general characterization of a domain can help to allay some worries about that domain, even if it is a substantive thesis within the domain itself and so does not address such external worries. In particular, a normative desire theory might support the idea of correctness, by characterizing reasons for action in a way that makes clear how it can be that questions about reasons have definite answers. It might even, via decision theory, allow for reasons about these answers to be given a formal structure.

Although it would be a mistake to reject a normative desire theory on the grounds I mentioned in Lecture 1 (lack of metahysical depth) such a theory would, like the Iterative Conception of Set, need to be justified by a larger reflective equilibrium argument: that is, by arguing that, on reflection, it provides the best explanation of the conclusions about reasons that seem to us most clearly correct. It seems to me to fail this justificatory test. One reason for rejecting it is that it seems to give clearly wrong answers in many cases. If I were to have a desire to eat my car, this in itself would give me no reason to try to do so. But this is not simply a case of rejecting a (perhaps appealing) general thesis on the strength of our intuitions about particular cases. Rather, as is often the case in reflective equilibrium arguments, reflection on these conflicts undermines the plausibility that the general thesis had to begin with.

As I pointed out in my first lecture, the appeal of a normative desire theory lies in part in the fact that the claims about reasons that it supports are grounded in something that is already true of the agent—his or her desire—thereby making a claim that the agent cannot deny without irrationality. Counterexamples of the kind just mentioned—in which a pointless desire seems clearly to provide no reason at all—call our attention to the inadequacy of such claims. They bring out the fact that the ability of a desire for q to provide a reason for an action that would promote q depends on the reasons for wanting q in the first place. This is an instance of the general divergence between claims about reasons and claims about rationality (that is to say, about what agents must treat as reasons on pain of irrationality.) If an agent sees something to be said for bringing about q, then it is irrational for her to deny that she has a reason to do what will promote q. But it does not follow that she has any reason to do this thing.

So the status of normative desire theory is more like that of Naïve Set Theory than like that of the Iterative Conception of Set. It does not lead to contradiction, but reflection on its counterintuitive implications leads us to see that its initial appeal was based on a mistake: a tendency to confuse conclusions about reasons with a certain kind of conclusion about rationality, and to fail to see that the reasons provided by a desire depend on reasons prior to that desire itself.

This leaves us with the question of what other general characterization of the domain of reasons there might be. Are we left simply with a diverse collection of intuitions about which things are practical reasons for action? The main candidates for this role seem to me to be these some form of constructivism or some other way of grounding reasons in an idea of rationality. I will try to explain why neither of these seems to me likely to succeed.

Broadly speaking, a constructivist account characterizes the facts about a subject matter by specifying some procedure through which these facts are determined, or "constructed." Many quite different accounts fit this broad definition. In order for an account to be constructivist in this broad sense, the procedures it specifies need not, as in mathematical constructivism of the kind Dummett argues for, be one though which we actually dertermine the truth values of statements in the target domain. This stricter version has obvious appeal, but I will allow weaker versions to count as constructivist in the sense I will discuss. So for example, a view might be constructivist if it held that the truths of a given domain are just those for which there is a construction (a proof or piece of reasoning) whether or not we know of it.

My Contractualist account of moral right and wrong would count as constructivist in this broad sense. It specifies that in order to determine whether an action is morally permissible we should consider a general principle that would permit it. We then consider what objections individuals might offer to this principle based on the way in which they would be affected by it: by living with the consequences of the actions it would permit and with the possibility that agents may perform such actions, since they would be permitted to do so. We then compare these reasons with the reasons that individuals would have to object to a principle that would forbid actions of the kind in question, based, again on how they would be affected by such a principle. We then compare these reasons, and consider whether it would be reasonable for those who have reason to object the principle permitting the action to reject it, given the reasons that others have for objecting to the contrary principle. If it would be reasonable to reject that principle, then the action is question would be morally wrong. The rightness or wrongness of an action depends on what the outcome of this procedure would be, whether or not anyone has carried it out.

This constructivist account of moral right and wrong depends on a prior understanding of reasons. The question we are concerned with is whether there might be a further constructivist account of reasons for action. One strategy for developing such an account would follow the model of the Iterative Conception of Set.

Sets are composed of their elements—other sets or individuals. This makes it very inviting to characterize the domain of sets by specifying how new sets can arise from others. The Iterative Conception of Set is one attempt to do this, and as that example illustrates, although such a view might be called “constructivist” since it specifies how the universe of sets is “constructed” it is not a form of constructivism in the sense in which this term is used in mathematics (the sense which Dummett uses), since this process of construction is not in general one that we can carry out, or see the results of. Following this model, however, it is an inviting strategy to characterize the domain of reasons by giving principles specifying how some reasons arise from others, or ways in which an agent has one reason in virtue of having others. Principles of means-ends reasoning (however they are to be specified) would be an example.

I need to pause here to note an important distinction between two ways in which a claim about reasons for action might be said to derive from or depend on other such claims. I myself want to defend a form of holism about practical reasoning: that conclusions about reasons for action are justified simply by thinking carefully about them in the mode described by the method of reflective equilibrium: considering what general principles about reasons would explain them, what implications these would have, considering the plausibility of the implications of these principles and so on. So, for example, suppose it seems to me that someone has reason to do a because he or she would find it pleasant. Does pleasure always constitute a reason? There are cases in which it does not seem to do so (taking pleasure in the pain of someone one dislikes, for example. How can we distinguish between pleasures that really do seem to provide reasons and those that do not? The conclusions we reach via this process will depend on clams we make along the way, and feel confidence in, about cases in which pleasure provides reasons and cases in which something plausibly called pleasure does not.

This kind of dependence, however, is a matter of how the justification for *believing* some claims about reasons for action depends on other such claims. The dependence involved in the kind of constructivism I am considering, the kind of which the principle of means/end reasoning is an example, is different: it is a matter of one consideration's *being* a reason for an agent because a certain other consideration is a reason for him or her.

A claim of dependence of this kind might be stated as a truth about reasons: there is sufficient reason to adopt end e, then the agent has a reason to do one of the actions a that would advance e.[10] Alternatively, as others would hold, it might be stated as a requirement of rationality: If one has adopted e as an end, then one must, insofar as one is not irrational, see the fact that would help to bring about e as a reason to do a.

Principles of both forms may be correct. (If the former truth about reasons holds, then one would expect fully rational thinking to reflect it in the way that the latter principle states: if a person believes that he or she has reason to promote e, then he or she will, if rational, believe that he or she has reason to do what promotes e. Even if this principle of rationality is correct, however, it is important to note that it yields no conclusions about what an agent has reason to do, but only conclusions about what agents must see as reasons, insofar as they are not irrational, given their other beliefs. This can be seen from the fact that this requirement applies even if the agent in fact has no reason at all to pursue e. The former, reasons-based version, makes this dependence on the reasons for e explicit by beginning, "If an agent has sufficient reason, …" This marks it,

[10] This is a version of, or very close to a version of, what Joseph Raz calls the Facilitative Principle in "The Myth of Instrumental Reasoning."

in my mind, as the more fundamental principle, even if some version of the other is also correct.

Let me call principles of this kind—of which principles underlying means-end reasoning are an example—principles of construction. How do we see that such principles are correct? I suggest—and I am going to take this as a working hypothesis—that they can be seen to be correct by reflection on the idea of what it is to be a reason for action (or alternatively, the idea of a rational agent.) I am not suggesting that these principles are analytic, or that they are "contained in" the concepts of a reason for action, or the concept of a rational agent. Rather, I am suggesting, they are seen as correct through a reflective equilibrium process of thinking about how the ideas of reasons for action, or rational agency, are best understood. (As you will no doubt guess, I am thinking here by analogy with the way in which axioms of set theory and the Iterative Conception of Set are arrived at by reflection on the concept of set.) In particular, the reflection that leads us to see that these principles of construction are correct does not rely on any judgments about which considerations are in fact reasons. I will say that claims that can be justified in this way have a *formal basis.*[11]

At this point we should note a structural difference between the realm of sets and that of reasons for action. A characterization of the domain of sets can be given entirely,

[11] I conjecture that the distinction between claims about reasons for action that have a formal basis and those that do not underlies the distinction Korsgaard draws between substantive and procedural realism. (*Sources of Normativity*, pp. ---) What she calls procedural realism relies (ultimately) only on claims about reasons for action that have a formal basis, substantive realism relies on some claims that do not. It is an interesting question, I think, whether all the modes of reasoning that constitute a "sound deliberative route" of the kind that Bernard Williams refers to have a formal basis. I suspect that the do not (that some steps in such a route depend on substantive claims about which tings are good reasons.) I argue for this in the Appendix to *What We Owe to Each Other*.

or almost entirely, in terms of principles of construction, which specify the existence of some sets given the existence of others (*almost* entirely if the existence of the empty set must be posited at the outset.) But a characterization of the realm of reasons for action cannot have the analogous form. In order for us to have reasons for action there must be some valid claims about reasons that do not, normatively, depend on other reasons in the way I described.[12] Indeed there are many such reasons, and a complete account of the epistemology of reasons for action must explain how we can come to know truths about these underived reasons for action. This is the most difficult part.

How do we come to know particular underived truths about which things are reasons? My own answer is that we proceed in the way I described above in discussing whether pleasure was a reason for action. We examine our responses carefully, consider what general claims about reasons they would lead to or be explained by, assess the plausibility of the implications of these more general claims in turn, seeking a reflective equilibrium. But this account faces a challenge. When we come to believe that p is a reason for us to do a, what supports our view that this is a case of coming to a correct conclusion about the reasons there are, rather than simply a reflection of our psychology, a quirk of some kind? Since the process of seeking reflective equilibrium depends at many points on our assessment of the plausibility of particular claims about reasons, why think that what it yields is anything other than a more or less accurate portrait of our particular psychological tendencies? (This epistemological challenge is the remnant of the question with which we began, about how we can "get in touch with" normative facts, freed now from its unnecessary metaphysical framework.)

[12] As Christine Korsgaard argues in "The Normativity of Instrumental Reason."

This challenge provides one line of thought leading to a Kantian view. This line of thought begins with the idea of a distinction between thinking that depends solely on the concept of a rational agent (or, I would say, of a reason for action) and thinking that involves focusing one's attention on some particular consideration, and finding it to be a reason for some action. All conclusions of the latter kind, Kant believed, depend on our subjective responses (our "inclinations".) Hence only the former can lead to conclusions that can be considered objective, because they have what I called above a formal basis.

Reflection of this formal sort, Kant believed, is constrained by two ideas. The first is that in order to see ourselves as agents, we must see our practical reasoning as having the capacity to asses and potentially overrule the appeal of any inclination or combination of inclinations. This means, he thought, that we must see the highest level principle of our practical reasoning as "purely formal"—that is, as appealing to the ideal of law-governed willing rather than appealing to the reason we have to promote any particular end. He thought there was only one such principle: the (universal law form of the) Categorical Imperative. The second constraint, according to some Kantians, is that we must see our wills as non-derivative sources of reasons—that is, as making ends valuable by choosing them.

Neither of these Kantian views seems to me tenable: neither his inclination-based interpretation of all our tendencies to see particular considerations as reasons, nor his interpretation of the conditions of rational agency. But even if one does not accept Kant's account of the two sides of this dichotomy, one might still accept the dichotomy and the line of thinking that it supports: that in order for us to have grounds for seeing conclusions about reasons for action as having any claim to objectivity we must see them as having a formal basis: as grounded in reflection on the idea of a reason, or of rational agency, rather than merely reflecting what we find attractive. For in the latter case, how can we tell that it is not just our subjectivity speaking?

This way of looking at the matter fills what otherwise seems to be a gap in Kant's discussion. In claiming that in order to see ourselves as acting, rather than merely being pushed around by our inclinations, we must see our highest level principle of practical reasoning as "merely formal" Kant seems to ignore the possibility that our ability to assess and potentially overrule our inclinations could lie in our ability to make judgments about what is good, or what we have reason to do. But how can an agent have adequate grounds for thinking that he or she is making a judgment of this kind rather than merely being attracted to a particular alternative for purely subjective reasons? This is a live question, which does not depend on distinctively Kantian premises. If the answer is that an agent can have such grounds only if he or she has ground for seeing the judgment as having what I have called a formal basis—as supported by the (the best interpretation of) very idea of a reason, or of rational agency—then this closes the gap just mentioned inKant's presentation of the argument.

This connects with, and complements, a point I made in my first lecture. I said there that part of the appeal of rationality-based views, for many people, is their ability to explain how reasons "get a grip on the agent." Such views enable us to go beyond merely saying that a consideration *is a reason for* the agent by linking that consideration to something already true of the agent, thereby making the reason one that the agent cannot deny without irrationality. I said that this explanation of the grip of reasons had greater significance in the context of interpersonal argument about reasons than it has from a first-person perspective. From that perspective, I said, the question that is relevant for an agent is just "*Is* this a reason?" The fact, if it is a fact, that the agent must, given the other things she takes to be reasons, see this consideration as a reason unless she is irrational, does not settle the matter. What matters is whether these other things really are reasons.

But when we focus, as we are now doing, on an agent's possible doubts about her answer to this substantive question, the "grip" that might be provided by a rationality based account may seem more relevant from the agent's own point of view. As I mentioned, Korsgaard says that from this point of view an agent must keep stepping back and asking "Why?" until it is "impossible, unnecessary or incoherent to ask why again." As I pointed out, the crucial question is when it is "unnecessary" to ask why again. This is, as I said, the substantive question of when the agent's confidence that some consideration is a reason is justified, and it is just this question that we are now confronting again. The Kantian view claims that an agent can have good ground for thinking that he has made a judgment about a reason only if this judgment can be grounded in the idea of a reason, or of rational agency. Otherwise, Korsgaard says, the agent has nothing to go on but his confidence that it is a reason, her suggestion being that this confidence will be misplaced.[13]

What conception of rationality, then, might play this role? I have explained why Kant's view does not seem to me satisfactory. Looking elsewhere, we should bear in mind that, as I pointed out in my first lecture, the relevant notion of rationality could not be the broad idea according to which what is rational is just what one has most reason to do. Appeal to that idea in this context would be circular. One alternative, I believe, is a

[13] Korsgaard, *Sources of Normativity*, p. ---

narrower notion of rationality, opposed to irrationality of the sort that involves holding incompatible attitudes, such as having an end but denying that the fact that a would promote this end is any reason to do a. This is the conception captured by requirements of rationality of the kind that John Broome discusses.[14] Principles of construction, for deriving some claims about reasons from others, are requirements of this kind. The question before us is whether they are the only such requirements, or whether being rational in this sense requires us to recognize some things as reasons. I agree with Broome in believing that it does not—that such requirements have only to do with the relation between an agent's attitudes.

Beyond this narrow idea of irrationality, Derek Parfit's suggests that there are some substantive claims about reasons that it is irrational to deny. He gives as an example thinking that it matters to one's reason for avoiding a pain whether it will occur on a Tuesday.[15] Perhaps this is correct, but I do not myself see any ground for calling this irrational other than the substantive obviousness of the claim in question. If this is correct, then such requirements will not have a formal basis.

I believe, although I cannot claim to have established this conclusively, that the claims about reasons that have a formal basis include only principles of construction, not substantive claims about what reasons we have. If this is correct, then the epistemological challenge that I have described cannot be answered through the quasi-Kantian strategy of appealing to judgments that have a formal basis.

The only method we have for arriving at and assessing particular conclusions about reasons is the one I gestured at in discussing the case of pleasure as a reason for

[14] Cite Broome papers.

[15] Parfit, ----

action: the method of seeking a reflective equilibrium of our substantive judgments about reasons for action. The question of whether a conclusion that we arrive at in this way is correct as a claim about reasons for action or a “quirk”—a mere manifestation of our particular psychology—is simply and only the first order normative question of whether the consideration in question really is a reason or not. It can only be answered by further reflection of this very same kind.

To see why we should not find this conclusion depressing, we should look back at the epistemological challenge and ask why it might have seemed that we would be in a stronger position respond to this challenge if we could show that our judgments about reasons had a “formal basis” in the best understanding of the concept of a reason for action, or of the concept of a rational agent. I can think of two possible advantages. One is that this basis would provide conclusions about reasons for action with normative authority—would give them a grip on the agent—by grounding them in something that an agent thinking about what reason he or she has for doing something already accepts. This advantage seems to me illusory. A claim about a reason for action, if correct, has the only kind of authority it needs. This is especially true from an agent’s own point of view, which is the one we are presently considering: if an agent accepts a judgment about a reason for action then he or she will see that reason as normative, and if not irrational, will respond to it accordingly. So I mention this possible advantage only to set it aside.

The second, properly epistemological advantage is that showing that a judgment about a reason for action had a formal basis in the concept of a reason or the concept of a rational agent would show that it was an answer to an intellectual question that, like a question about sets, has a determinate “objective” answer, an answer that we can discover using familiar intellectual abilities to reflect on our concepts and work out the best interpretation of them. But we should recall here that although the axioms of set theory, the Iterative Conception of Set, and claims about the requirements of rationality are supported by reflection of the relevant concepts, none of these things is analytic—their support lies rather in their being the outcome of reflective equilibrium-seeking processes aimed at finding the best way to understand these ideas. (This is, in particular, the most, or I would say more than the most, that can be claimed for Kant’s conclusions about the conditions of rational agency.) The claim of these conclusions to objectivity depends simply on the quality and authority of this process. Substantive conclusions about reasons that are not formally based seem, by contrast, “subjective” only if they are assumed to be isolated individual responses, like occurrences of a desire. But the judgments about reasons that survive the kind of reflective equilibrium process I have described are not like this. They, too have undergone careful reflection and reexamination. Perhaps it will be said that the process of reflection through which we arrive at an overall view of reasons for action is not an *intellectual* process in the relevant sense. But this seems to me a mere prejudice.

The determinateness of claims about reasons—the degree to which such claims have definite truth values—depends on the outcome of this process. I will explore these matters further in my final lecture.