24 *Reflection*

Classical rationalists appealed to a capacity for reflection. As a first approximation, reflection is a type of rational cognition with four significant features. It goes beyond what is immediately obvious. It is higher-order, in the sense that it involves thought about psychological states or representational contents, although its conclusions need not be about psychological states or representational contents. It aims at constitutive understanding. And it develops such understanding by drawing conclusions, without acquiring new premises, empirical or otherwise, beyond what is already understood or known. Usually reflection aims at improving pre-reflective understanding or knowledge. Much reflection makes use of empirical background knowledge. Where the force of its warrant is independent of sense experience, reflection is apriori.

Some apriori knowledge needs no reflection. Knowledge that $2 + 2 = 4$, that cruelty is wrong, and that everything is self-identical seem both too obvious and too first-order to be a product of reflection. Apriori knowledge deriving from first-order reasoning from first-order knowledge also does not derive from reflection. Thus much reasoning in mathematics does not.

Traditionally, reflection was supposed to begin with an attempt to clarify ideas, concepts, or relations among cognitive capacities. It was supposed to arrive at truths by considering thoughts, or principles guiding thoughts—even if the ultimate subject matter was not the thoughts themselves. Thus reflection has played at least an auxiliary role in any both empirical and mathematical sciences, as well as in philosophy.

Obviousness is perhaps a matter of degree. Involvement of a higher-order component is perhaps a matter of emphasis. Clarifying concepts or meanings and development of first-order theory about a subject matter are interdependent. I will not try to sharpen these matters here.

A lot of apriori knowledge, however, seems to utilize and require reflection. Philosophy offers examples of reflection's yielding purported apriori knowledge or apriori warranted belief. Philosophy is not alone. Meta-logic offers examples of reflection's yielding apriori knowledge.[1] Neither of these disciplines invokes

[1] Kant assimilated most or all reflection to reasoning in either general logic or what he called 'transcendental logic'. *Critique of Pure Reason* A261–263/B316–319.

special cognitive powers. They use refinements of forms of understanding and reasoning that are available to common sense and to first-order theorizing.

Although the notion of reflection is only roughly delimited and is incompletely understood by all of us, I take reflection to be a source of knowledge and warranted belief. Some of the knowledge and warrant are, I think, apriori.

I

The classical rationalist conception of reflection holds three main theses. The first is that *in reflection an individual always brings to articulated consciousness steps or conclusions that are implicitly present, subliminally or unconsciously, in the individual's mind before reflection.*

The relevant sense of 'implicit' is *unconscious but capable of being accessed to consciousness through reflection*.[2] The idea is that the material that is accessed through reflection not only marks a mental state that was present before reflection. The mental state so marked must itself be accessible to consciousness through reflection. Reflection brings to full consciousness thought contents and attitudes that were already harbored, though only unconsciously or dimly.

All major classical rationalists thought that all materials for understanding the principles obtained through reflection are present in the mind before and during every stage of reflection.[3] This view was often motivated by the idea that in thought and reasoning, we "must know what we are doing": We must implicitly understand the principles that govern our thoughts. It is these principles that we get at when we succeed in making them explicit through reflection.

The second thesis is that *reflection can yield apriori knowledge of objective subject matters, beyond thoughts that the reflector is engaging in.*

'Apriori' means independent of sense experience for its warranting force. 'Objectivity' carries a variety of senses, depending on the author. A minimum sense is independent of the idiosyncrasies of the individual.

The classical rationalists thought that reflection yields knowledge of subject matters that are not explicitly mental. They thought that even with regard to mind, reflection yields knowledge or cognition of principles that are necessary and universal. So they thought that some of the apriori knowledge applies beyond the

[2] Leibniz, *New Essays on Human Understanding*, Preface, 51, provides a closely related characterization: '. . . reflection is nothing but attention to what is within us . . . '.

[3] Descartes, *Meditations*, V; Leibniz, *New Essays on Human Understanding*, I, i, 73ff.; Kant, *Critique of Pure Reason* A261–263/B316–319. Descartes thought that reflection consists in clear and distinct cognition of natures and their relations. Leibniz thought that it consists in analysis of the structure of ideas or concepts. Kant thought that Leibniz was right about reflection in logic, but that the most fruitful reflection derived from bringing to consciousness a subliminal self-awareness of the procedures governing use of ideas or concepts. Kant went so far as to hold that reflection itself is present, though implicit, in the natural pre-theoretic operations of all thought. See *Critique of Pure Reason*, A103–108; *Anthropology* section 7, AA vii, 141.

psychology of the particular individual doing the reflecting. The thesis entails that reflection can yield knowledge that is distinct from the self-knowledge involved in awareness of what thoughts one is thinking.

Again, different writers explain these epistemic commitments in different ways. Descartes maintains simply that what one reflects upon is relations among ideas, and that these relations include self-evident principles. He holds that the principles emerge when one makes the ideas clear and distinct. But he provides no satisfying account of how one goes about making ideas clear and distinct. For this deficiency he was criticized by all his successors.

Leibniz develops Descartes' views in a particular way: all truths are truths of analytic containment of one idea or concept in another. So for Leibniz reflection takes the form of analyzing ideas or concepts to find definitional relations among them. The definitional relations are taken to reflect an order of epistemic priority.[4]

Kant takes Leibniz's view of reflection to be roughly correct for general logic, but incorrect beyond general logic. He maintains that general logic yields no substantive cognition of a subject matter—but merely presents rules and principles for rational thinking. So for reflective knowledge or cognition that is substantial, he proposes a significantly different picture of what reflection is up to. He holds that reflection can yield cognition of a subject matter only by giving up analysis of the content of thoughts or concepts and by centering on the function and use of mental capacities. He holds that reflection yields cognition of the "form" of such capacities, which he takes to include principles and rules governing their correct use.[5]

All these authors emphasize the normative relevance of principles obtained through reflection. Such principles yield norms for thinking as one "ought"—and not just because the principles are true. The principles are supposed to provide a framework for all of thought. They provide norms for the right direction of the mind in pre-theoretic and unreflective theoretical domains.

Descartes and Leibniz rest huge metaphysical systems on reflection. As is well known, Kant criticized these systems and metaphysical inquiry in general. He gives reflection the role of yielding apriori principles governing the form of cognition. But he agrees with his predecessors in regarding reflection as a source of apriori cognition of necessities with universal application.

The third thesis is that *successful reflection requires skillful reasoning and is difficult: it is not a matter of one-off introspection or intuition*. Descartes thought

[4] For further discussion of these matters, see my 'Frege on Apriority', in C. Peacocke and P. Boghossian (eds.), *New Essays on the A Priori* (Oxford: Oxford University Press, 2000); and 'Frege on Knowing the Foundation', *Mind* 107 (1998), 305–347. Both are reprinted in my *Truth, Thought, Reason: Essays on Frege* (Oxford: Clarendon Press, 2005).

[5] Two further doctrines that are very particular to Kant are (a) that cognition of form falls short of cognition of actual entities that instantiate the form; and (b) that form is made by the mind, not discovered as a mind-independent subject matter. The former view is part of his rejection of traditional metaphysics in favor of a "transcendental" meta-theory about possible cognition. The latter view is fundamental to his idealism.

that most people never come to understand any of their own ideas. Perhaps he thought that most people never reflect. But equally he thought that successful reflection requires great skill.[6]

Classical rationalists hold that even though reflection often yields objectively self-evident principles (their warrant lies in themselves), reflective knowledge is never obvious to just anyone who tries to reflect. One begins with incomplete understanding of what one reflects on, at the conscious, articulate level of thinking. Reflection is held to require following a method. Often the method is dialectical.[7] Reflective method involves following rational principles in a systematic, discursive, often dialectical manner.[8]

II

I accept the second and third theses and reject the first.

The third thesis seems to me exactly right. The general warning that reflection is not to be conflated with simple introspection or with spontaneous judgment is well taken. Reflection follows discursive methods. Some people are better at it than others.

I also accept the second thesis. I think that reflection has been a source of apriori knowledge that is not confined to self-knowledge. Of course, I differ with classical rationalists about the scope of reflection for yielding apriori knowledge. Nearly all philosophers since Kant do. Leibniz and Descartes are wildly overoptimistic regarding how much reflection can tell us about the nature of the physical world, and even of mind. I believe that both philosophers seriously underestimate the role of empirical experiment in natural and psychological science. Kant, who I think understood natural science much better, still overestimates the role of reflection in determining principles for natural science; and he overestimates the scope of reflection in certain other ways. On the other hand, I believe that Kant underestimated the scope for reflection in both logic and philosophy of mind.

I think that Descartes and Leibniz are right to hold that some not immediately obvious principles can be understood to be self-evident through reflection. Leibniz's appeal to analysis of concepts also has a genuine but extremely limited

[6] Descartes, *Rules for the Direction of the Mind*, Rule IX, XII. See also Leibniz, *New Essays*, I, i, sections 23–27, 85–88. Kant, *Critique of Pure Reason*, B1–2. The *hubris* of classical and even Kantian rationalism was slightly mitigated by an emphasis on the difficulty of following rationalist methods, and the need for following the direction of the master teacher.

[7] The Platonic dialogs provided a historical precedent for elaborating the dialectical methodology of reflection.

[8] Descartes and Leibniz favored the Platonic metaphor of vision for describing the understanding arrived at through successful reflection. Although Kant occasionally uses the visual metaphor (*Critique of Pure Reason* A108), he is more like Frege and twentieth-century authors in emphasizing the essentially discursive character of understanding. In *practice*, even Plato, Descartes, and Leibniz reflect much more discursively than their vision-dominated glosses suggest.

role to play in reflection. Kant's centering reflection on the function and use of mental capacities seems to me a fundamental step forward.

Some of the methods classical rationalists used to achieve apriori knowledge through reflection now seem naive. But I believe that they were right to strive for apriori knowledge, and sometimes successful in finding it. Moreover, some special normative status for principles obtained from reflection also seems an inevitable consequence of the second thesis.

My central disagreement with the classical rationalist conception of reflection centers on the first thesis. I will summarize, as background, points I have made in earlier work.

The first thesis maintains that reflecting individuals always bring to consciousness, through discursive methods, what is already implicitly present in their own minds. Anti-individualism has combined with advances in psychology to undermine this thesis.

Anti-individualism is the view that the nature of many representational mental states is constitutively dependent on relations between the individual who has those states and aspects of the environment, or other objective subject matter. The relations can be complex—running through other mental states or through other individuals, including the individual's species-ancestors. The key point is that many mental states are constitutively dependent on there being non-representational, explanatorily relevant relations between the individual and the environment or other subject matter. I have argued for this view elsewhere.[9] The upshot of this view is that the representational states one is in are less a matter of cognitive control and internal mastery, even "implicit" cognitive control and mastery, than classical views assumed.[10]

Two main advances in psychology also bear on the first thesis. One is the fact that capacities of certain animals and young children are best explained in terms of simple propositional attitudes. Since animals and young children seemingly lack the conceptual competence to make their own attitudes the topic of their thinking, they seem to have first-order thought while lacking a capacity for the higher-order thought involved in reflection. Nearly all classical rationalists maintain, on I think no good grounds, that animals lack propositional attitudes. Individuals can think without "knowing what they are doing". They can even lack a capacity to develop a propositional understanding of their thinking. An individual need not be able to represent principles or even inference rules that govern the individual's thinking. Principles and conceptual relations may explain mental activities without guiding them—without the individual's following them

[9] [Added 2011:] See my *Foundations of Mind: Philosophical Essays, Volume II* (Oxford: Oxford University Press, 2007).

[10] This is not to say that the rationalists rejected anti-individualism. I think that Descartes, for example, accepted it. See my 'Descartes on Anti-Individualism' in *Foundations of Mind*. It is more that they did not think through the implications of anti-individualism.

by implicitly knowing them. Not all cognitive activity that is correctly explained as following rules involves a capacity to represent the rules, even unconsciously.

The other advance in psychology is recognition that many principles governing mental activity, even in adult human beings, are modular, or otherwise inaccessible to being reflectively brought to consciousness. Even principles governing non-modular, accessible cognitive activity are often inaccessible in this way.

These facts refute Leibniz's and Kant's extreme claim that all thought constitutively involves reflection. But they do not in themselves refute the first thesis. One could hold that *in those domains, and in those types of reasoners, where successful reflection is possible*, the mental state marked by the relevant principle is in the mind. But such a position is not plausible in light of both psychology and the history of science. Psychology has brought out that accessible higher-order, person-level cognitive control plays a very small role in much of our propositional activity. Principles that best explicate a mathematical or natural-scientific concept are often discoverable only by developing new knowledge, knowledge that it is not psychologically plausible to impute to the unconscious of reasoners before the new knowledge is discovered. Sometimes explication of principles requires development of new concepts. Reflection on one's past usage can yield principles that were not present in the mind before the reflection began. Reflection seems sometimes to produce understanding of a principle that did not mark any antecedent mental state.

Let us develop these considerations at somewhat closer range. In the twentieth century a number of philosophers have emphasized the role of *examples* in attaining illumination through reflection. This emphasis is not new. Some classical rationalists, particularly Socrates and Descartes, used examples prominently in reflection. The aim is to use examples to arrive at principles. The classical idea is that in making judgements about examples, we are guided by principles. The examples help make the principles more explicit. I believe that this idea is sometimes correct. But its scope must be qualified in light of the considerations just advanced.

There is reason to believe that principles that explain judgments about examples are not always stored in the mind of the judge. The person may go on perceptual similarity, where the general principles governing similarity are not yet conceptualized. The principles can explain the interaction of perception and perceptual memory, but be nowhere represented in the individual's psychology. For example, in judging something to be a chair, an individual might operate under principles governing similarity to a set of perceptual paradigms—but might lack conceptualization of the principles. Or, for other reasons, an individual might never have formed a propositional attitude toward the principles, however implicitly.[11]

[11] See my 'Intellectual Norms and Foundations of Mind', *The Journal of Philosophy* 83 (1986), 697–720; reprinted in *Foundations of Mind*; 'Concepts, Conceptions, Reflective Understanding', in

This kind of point applies in non-empirical domains. A mathematician might operate with a sense that two examples are similar enough to fall under a single concept, but lack the concepts necessary to make systematic inferential use of the principles governing the sense of similarity. The concepts in such principles may not yet be formed by the individual. The principles may, for other reasons, not be internal to the individual's psychology. Sub-conceptual senses of similarity may be the psychological basis for the judgments about examples. Conceptualization of the principles that cover the cases may come later.

Interlocution also plays a role in explaining judgments about examples. An individual can make judgments through memory of being told that the example falls under some concept. One might judge a seat in a ski-lift to be a chair because one remembers its being counted a chair—without having formed any general principle that unites the examples that would be counted chairs. Again, what explains and vindicates the individual's application of a concept to cases may not be a conceptualized principle internal to the individual's psychology.[12]

In some cases, the relevant principles governing *correct* application transcend what is available even to a community. In the case of natural kinds, there seems to be some conditional commitment in individuals' psychology to the effect that *if* a concept is a natural kind concept (something that cannot be known apriori), it applies to all instances of the principles that in fact govern the natural kind. These principles may not yet be discovered. The psychological ability to track examples lies in an ability to track superficial properties of the examples that by-and-large line up with the natural kinds in question. I will mention analogous cases in mathematics later.

More broadly, in a wide range of judgments about examples, there is a commitment in thinkers to the objectivity of the subject matter, whether or not the relevant entities instantiate kinds that interest science or mathematics. The correct explication of principles governing the application of concepts can depend on the outcome of investigations that the individual has not engaged in. Even experts may not have internalized principles that explain correct application.[13]

This latter point is supported by consideration not only of thought experiments but also of intellectual history. The correct account of the application conditions for concepts may emerge long after those concepts are in use. This sort of situation is most vividly illustrated in the history of scientific or mathematical concepts. But it is potentially applicable over a wider range.[14] Sometimes, one

M. Hahn and B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.: MIT Press, 2003); reprinted in this volume.

[12] There is a tendency in classical rationalism to underestimate the role of interlocution and reliance on others in both common sense and scientific domains.

[13] 'Intellectual Norms and Foundations of Mind'.

[14] See 'Intellectual Norms and Foundations of Mind' and 'Frege on Sense and Linguistic Meaning', in D. Bell and N. Cooper (eds.), *The Analytic Tradition: Meaning, Thought, and Knowledge* (Oxford: Blackwell, 1990); reprinted in *Truth, Thought, Reason.*

can make progress in understanding a concept and the phenomonea that it applies to by reflecting on applications that had previously been held together either by a non-conceptualized sense of similarity, or by acceptance of the word of others.

I and others have emphasized that individuals can harbor mistaken explicatory beliefs.[15] Even settled views about principles that govern the application of concepts, purportedly providing necessary and/or sufficient conditions for falling under a concept, are frequently mistaken. In itself, the classical view of reflection accepts such cases. By the third thesis, it emphasizes the difficulty of finding correct principles. It must, however, maintain that the individual's incorrect explications are incompatible with correct principles that are, according to the first thesis, "implicitly" stored in the mind.

Both psychology and anti-individualism have cast incorrect explications in a different light. The correct principles need not be stored in the individual's psychology at all. Explanation of the psychology of incorrect explicatory judgments about examples often should not appeal to implicit internal contradictions. Judgments about examples are often correctly explained without assuming that principles governing the psychology or the correctness of the judgments are the content of some prior psychological state. Yet reflection can sometimes unify what was not previously held together by a principle unconsciously represented.

These various grounds for thinking that explicatory principles that explain correct judgments about examples need not be internalized support a different view of concepts and conceptual psychological ability than that which guided classical views of reflection.

Leibniz maintained that reflection consists in analyzing concepts so as to elicit principles contained within them. Developments in philosophy and psychology have long conspired to undermine the view of Leibniz that most concepts are definable in terms of other concepts, and that having concepts is being in psychological states whose content consists in the definitional relations, which, in turn, constitute principles for the correct application of the defined concepts.

Kant maintained that most concepts are constituted out of a complex conceptual structure contained within them. But as noted earlier, he put little weight on analysis of concepts in his account of the role of reflection in obtaining philosophically interesting results.[16] Like Leibniz, Kant tended to identify concepts with the "function" or rule for its correct application, though he thought that this conceptual structure does not in general suffice to provide sufficient conditions for correct application of concepts.[17] Sensible intuition is often needed over and above concepts. I do not accept Kant's claim that all substantive theoretical cognition requires sensible intuition (perception or singular representations of aspects of the structure of perception) for its warrant. For example, I think that

[15] See Hilary Putnam, 'Is Semantics Possible?' in his *Philosophical Papers*, volume II: *Mind, Language and Reality* (Cambridge: Cambridge University Press, 1975).

[16] See *Critique of Pure Reason* A65–66/B90.

[17] *Critique of Pure Reason* A68–69/B92–94.

cognition in mathematics does not rest on *sensible* intuition.[18] But I think that Kant's invocation of intuition is a major step toward recognizing the centrality and irreducibility of examples in the determination of concepts, hence in reflection.

Kant differed from Leibniz in that he thought that the way to come to understand relevant rules through reflection is not conceptual analysis, but consideration of the inferential and applicational (judgment) capacities marked by the concept, and its relations to sensible intuition. He thought that the relevant principles are fully conceptualized—even though he supposed that the most important ones are meta-rules that connect concepts and principles to sensible intuition. He retained Leibniz's view that having a conceptual ability consists in, or is necessarily associated with, being in a psychological state that has the content of the rule or principle that guides and explains the concept's correct application.

What is the alternative view of conceptual psychological ability? The view is that one can have a conceptual ability to think in a certain way about objective matters, even though the principles explaining that ability are not fully internalized. The principles can involve concepts not available to the individual with the relevant conceptual ability. They may govern not only the thinker's thinking but his perceptual abilities or his sense of mathematical similarity. They may include reference to communicative relations and reliance on the abilities of others. They may involve reference to patterns in the environmental or abstract subject matter that the individual thinks about. The individual's incomplete understanding of the concept—his inability to formulate the principles governing his ability to apply it correctly—can go beyond an inability to make explicit the content of subliminal, but still accessible, psychological states. Thought and the having of thought content are not in general to be explained purely internally. Conceptual ability is not individualistically self-contained.

These considerations show reflection to be more difficult than classical rationalists thought it was. Reflection cannot always expect to find true principles governing pre-theoretic thinking, or unreflective thinking in scientific theory, already present as the content of a psychological state. To obtain answers it sometimes must await or develop materials that go beyond what is available in the individual's current psychology. One way to do this is to use empirical knowledge. Another is to engage in first-order apriori theorizing, perhaps thereby learning new things or even producing new concepts. Even though I think that sometimes an individual can in reflection develop new concepts and principles, it seems clear that, to be successful, reflection sometimes depends on incorporation of discoveries that give it the tools to provide retrospective explications or principles governing a given way of thinking. To obtain answers, it cannot always simply bring to consciousness what is already implicitly known or represented.

[18] See 'Frege on Apriority'.

The same considerations indicate that reflection has a more limited role in acquiring knowledge than the classical rationalists supposed. Since reflection is not guaranteed a subject matter complete in itself, reflection must be given a secondary role. Much knowledge, including apriori knowledge, is obtained without heavy reliance on higher-order consideration of thoughts or psychological states. First-order theorizing is dominant in our acquisition of knowledge.

For all that, I believe that reflection is sometimes a source of knowledge. So far, most of the points that I have made about reflection have been limitative. I would like to make some progress in understanding reflection as a genuine source of knowledge by considering cases. I begin with a brief review of the success of reflection in meta-logic.

III

Reflection produced spectacular, sustained success in the development of meta-logic.[19] I want to consider some salient aspects of the reflection, extended over a long history, that led to the elaboration of certain intuitive concepts as the basis for modern model theory.

The intuitive notions are logical consequence and logical validity. These notions are a product of reflection on good deductive inference. The intuitive notion of *logical consequence* is that of *preservation of truth grounded in or correctly explicable in terms of logical form and logical structure*. *Logical validity* is simply *truth grounded in or correctly explicable in terms of logical form and logical structure*. Here I distinguish form and structure. Form is a property of sentences and propositional thought contents. Structure is the aspect of the *subject matter* of sentences or thought contents that helps explain systematic preservation of truth.

Let me summarize the development of reflection that issued in these notions and their productive use in leading to modern meta-logic. The oldest discussions of deductive inference center on *necessary connections* between the truth of premises and the truth of conclusions, and on unexplicated notions of demonstratively derived *knowledge*. The intuitive notion of logical consequence emerges in the middle ages. This emergence required isolating truth preservation from knowledge preservation, and separating out a conception of form from the modal notion of necessity as the distinguishing mark of good deductive inference. The point here is not that there was any rejection of modal or epistemic notions. All the relevant writers assumed that in good deductive inference, there is a necessary connection between premises and conclusion. All assumed that

[19] In 'Logic and Analyticity', *Grazer Philosophische Studien* 66 (2003), 199–249, I have traced the emergence of the intuitive concepts of logical consequence and logical validity that became the basis for mainstream modern meta-theory in logic. I will be drawing on points in that article, where the points are developed in greater detail.

argument serves knowledge. Reflection isolated logical consequence and logical validity as phenomena in their own right.

This development included both concept formation and theory building. Yet it derived from reflection on instances of good deductive inference.[20] Reflection on good deductive inference yielded insight into the function of inference, the form of inferences, the type of concept that truth is, and the potential of those forms and their semantical relations to explain good deductive inference.

Reflection on *good* deductive inference is reflection on inference that fulfills a function, aim, or goal. What is the function of deductive inference? There is no single answer, since there are different sorts of function. There are surely social, biological, psychological functions of inference. These are not best understood through reflection. There are, however, other functions that can be understood reflectively. Deductive inference serves belief and judgment. Belief and judgment have a representational function of presenting an individual with true thoughts. Deductive inference functions representationally to preserve truth in argument. The foregoing train of reasoning is part of what underlay development of the intuitive notion of logical consequence.

Three more insights were needed. One was reflection on specific forms of sentences, propositions, or thought contents, in the light of how patterns of good deductive inference are associated with patterns among the forms. A second was the idea that since truth is some sort of fitting with a subject matter—since truth is a semantical notion—deductive inference is associated with patterns not only among the forms but among relations between the forms and structural aspects of a subject matter. The third is that good deductive inference can be explained in terms of patterns of relations between logically relevant forms and structures of subject matters. The flowering of model theory derived from Frege's making these four insights systematic, and Skolem, Gödel, and Tarski's mathematicizing them.

These insights are unquestionably successes for reflection. They may provide clues to finding aspects of reflection that yield success in other areas—in particular, philosophy.

All four insights require the prior recognition of good deductive inference. 'Deductive' and 'inference' need not have been understood precisely. It was enough to recognize a range of cases that serve as paradigms and that hang together. What is striking is the self-evidence of the first-order deductive inferences. This self-evidence supports the meta-characterization of them as good, as fulfilling a representational function.

[20] What is the significance of isolating the key intuitive notions from their metaphysical and epistemic cousins? Necessity was a more metaphysical notion than was needed in logical theory. The epistemic cousin of logical consequence is conceptually posterior and more complex. The simpler, more basic notions are truth and preservation of truth.

The first insight is *into a function of a representational activity*. Reflection on function is closely associated with reflection on standards or norms for fulfilling the function. Recognizing preservation of truth as a relevant good, and relevant function, of inference is closely associated with recognizing what it takes to fulfill that function in deductive inference. This is not a matter of degree. Either one reasons according rules of correct deductive inference or one does not. Still, it is one thing to know what the function is and another to recognize good ways to fulfill it. The relevant function is a certain type of preservation of truth—preservation of truth in virtue of form—a form that potentially reveals structure in the subject matters of inference.

The second insight is *into the form of a representational activity*. Insight into form is also insight *into relations among capacities*. For representational contents, including their forms, help type-individuate representational capacities. On reflection we recognize that the capacity to think that all humans are mortal shares a capacity with the thought that all dogs are mammals.

The third insight is into the *type of concept* that truth is, and the *type of property* that truth is. This insight involves recognizing necessary relations among truth, reference, and truth-of. This is an insight into the domain of application of a concept. Those who came to employ the intuitive concept of logical consequence recognized that truth in general is not just a matter of form. True thoughts characterize a subject matter, and thus bear a relation to it. So the notion of truth is to be understood in conjunction with notions of relation between parts of sentences or thought contents and a subject matter. So deductive preservation of truth is to be understood not only in terms of form (or relations among abilities) but in terms of the reference or the truth-of relations that parts of sentences or thought contents (or the underlying abilities) bear to a subject matter.

The fourth insight is into the *possibility of better understanding one sort of phenomenon in terms of another*. Good deductive inference (the relevant sort of truth preservation) is to be understood in terms of truth preservation grounded in logical form and logical structure. The reflective insight is that the intuitive notion of good deductive inference is to be systematically understood in terms of another intuitive notion, logical consequence, whose explanatory potential is richer. Eventually, the latter notion received systematization through use of the mathematically more precise notions truth of and truth in a model.

All of these insights derived at least partly from reflection. All are apriori. Not all were beyond question. One could rationally question, until the early twentieth century, whether good deductive inference was illuminatingly understood in terms of form and structure—independently of modal notions. I think it now safe to take these insights as successes of apriori reflection. I hope to juxtapose these cases with cases of successful reflection in philosophy.

IV

I want to discuss two examples of successful philosophical reflection.

The first underlies the principle that error presupposes a background of veridicality.[21]

The representational function of belief and perception is to represent veridically. A belief is representationally successful when it is true. A perception is representationally successful when it is correct or accurate. When an individual has a certain belief or perception, the conditions for representational success are set. The belief or perception is partly individuated in terms of its representational content: What it is to be a belief or perception of a particular kind is partly to have the content it has. A representational content is a condition for veridical, successful, representation. For belief, the content is a condition that when met is true. For perception, the content is a condition that when met is correct or accurate. So belief and perception are partly individuated in terms of conditions for their representational success. These are conditions under which their representational functions are fulfilled.

The key idea behind the principle that error presupposes a background of veridicality is that cases in which certain representational successes obtain are in some way more basic than cases under which errors, or representational failures, obtain. Part of the idea is that failure can be understood only as failure at doing something, and the "doing something" constitutes conditions of success. So the notion of failure is given its meaning by conditions for success. This is a relatively easy conceptual point.

But the key idea is stronger. The principle suggests constitutive explanatory connection between *cases* of success and *cases* of failure. The possibility of failure (error) is to be explained in terms of actual conditions that make representational success, cases of veridicality, possible.

I believe that a deeper account of why the principle is true brings in reflection on anti-individualism. But here I mention two initial considerations that support the principle.

One is that the mental states' having representational content is incomprehensible apart from capacities to use the content—associate it with sensory input,

[21] The principle—and its cousin, 'irrationality presupposes a background of rationality'—have antecedents in various formulations of a principle of charity. In *Word and Object* this principle was formulated: '. . . assertions startingly false on the face of them are likely to turn on hidden differences of language'. Quine added, 'The common sense behind the maxim is that one's interlocutor's silliness, beyond a certain point, is less likely than bad translation—or, in the domestic case, linguistic divergence.' *Word and Object* (Cambridge, Mass., MIT Press, 1960), 59. Davidson formulated a similar view: 'Making sense of the utterances and behaviour of others, even their most aberrant behaviour, requires us to find a great deal of reason and truth in them.' 'Belief and the Basis of Meaning' in *Inquiries into Truth and Interpretation*, 2nd edition (Oxford: Oxford University Press, 2001), 153; and '. . . errors take their content from a background of veridical thought and honest assertion', 'The Irreducibility of the Concept of Self ', in *Subjective, Intersubjective, Objective* (Oxford: Oxford University Press, 2001), 89.

apply it, use it in action, make inferences with it, and so on. Since the representational content makes reference to a subject matter, the capacities to use it must associate the content to the subject matter that it is about. Erroneous applications would not *per se* relate the content to the subject matter that it is about.[22] So in explaining the having of content, one must advert somewhere to capacities to get things right. Not every such capacity need be realized in every individual. But mental capacities can be attributed only against a background of some actual realizations. The states are part of a system for grounding explanation of actual mental activity. So in the constitutive explanation of errors, there must lie somewhere in the background an appeal to veridical realizations of mental capacities.

A secondary consideration is the long intellectual experience of expecting functions to be either explained by or essentially associated with patterns of actual acts or events. Although I think that not all functions are like this, I think that *specific* mental states' function to represent veridically, involving with *specific* veridicality conditions, must be explanatorily associated with actual acts or events. Such functions are not well understood as brutely inherent in states, acts, or events, completely abstracted from patterns in which they are grounded. The point here is similar to the point that the function of a heart cannot be individuatively understood as an isolated brute fact. One must relate it explanatorily to actual patterns of events that successfully realize needs of the body. Thus having a function of representing a particular subject matter is grounded not only in a network of further relations to the subject matter, but in some pattern of acts or events that successfully realize the function. Only in such a context are failures possible.

Making the error-presupposes-veridicality principle more specific is difficult. How much error and what sort of error are possible? What sort of explanation or grounding—and at what level of specificity—is to be expected? These difficulties are coded in the vagueness of the words 'presupposes' and 'background'—in 'error presupposes a background of veridicality'.

I believe that philosophy is still in the process of reflecting on more specific formulations. Here reflection proceeds dialectically. Consider one of Davidson's formulations: Most of an individual's beliefs have to be true.[23] It is natural to think of possible counterexamples, perhaps even actual ones. We seem to be able coherently to imagine an individual whose every empirical judgment is mistaken—perhaps the result of being put in an extremely abnormal situation. Of course, such counterexamples themselves might involve a mistake. There is scope here for good or bad philosophical judgment. I will, however, take this counterexample as sound. The case forces one to think of the principle in a wider

[22] Veridical aspects of erroneous applications could relate content to subject matter, of course. One could perceive a grey rabbit as a brown rabbit.

[23] Donald Davidson, 'A Coherence Theory of Truth and Knowledge', in E. LePore (ed.), *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson* (Oxford: Basil Blackwell, 1986), 308, 314.

context. The individual's errors are possible only against a wider background of successes. The individual's perceptual system could not represent the physical shapes that it represents (while being fooled) if the perceptual system had not evolved or been designed in an environment in which the same types of perceptual states in other individuals responded to actual physical shapes. Take another counterexample, there are areas of thought—say, astrology—where error is dominant. What is to prevent an individual's harboring more beliefs about astrology than about anything else? Reflection suggests that certain types of representational success, perhaps perception or simple logical inference, are central to understanding other successes—and the failures.[24]

Progress in sharpening principles such as 'error presupposes a background of veridicality' is slow. Often progress does not issue in scientific theory. But sometimes we do make progress that deepens understanding and yields knowledge, even where system and rigor never emerge.

Let me turn to a second example of successful philosophical reflection—the thought experiments that underlie anti-individualism. These thought experiments proceed in three stages.

The thought experiments begin with a judgment that someone could have thoughts about a given kind or property as such, even though that person is not omniscient about its nature.[25] For example, one can think that arthritis is a painful disease, or that aluminum is a light metal. Or someone could correctly deny that sofas are really religious icons. One could do these things even though one might lack the background knowledge to distinguish those kinds from imposters. One might not realize that arthritis can occur only in joints, or that aluminum has the atomic weight that distinguishes it from other metals that are perceptually and practically similar. One might lack the sophistication to counter a sceptic who incorrectly claims that sofas are not used for sitting, but are religious icons for a sect that fools the population into thinking that sofas are used for sitting.

Then, in the second stage, one imagines a situation like the first one except that a kind or property that is indiscernible (for the individual) from the original kind or property occurs systematically in its place. One also imagines that there is no social network that the individual relies upon that uniquely specifies the kind or property or that interacts with it. For example, one imagines an environment in which no one has isolated arthritis as a disease. Or one imagines an environment in which there is no aluminum, or no sofas. No one in the imagined environment can uniquely specify these entities. In their places are, respectively, another disease that feels about the same, another metal that looks and behaves about the same, and

[24] I believe *de re* representational states are fundamental.

[25] The thought experiments are laid out in detail in 'Individualism and the Mental', *Midwest Studies in Philosophy* 4 (1979), 73–121; 'Other Bodies', in A. Woodfield (ed.), *Thought and Object* (Oxford: Clarendon Press, 1982); and 'Intellectual Norms and Foundations of Mind'; all are reprinted in *Foundations of Mind*. See also the discussion of the thought experiments in that volume's Introduction and in 'Postscript: "Individualism and the Mental"'.

religious icons like those that the sophisticated sceptic mistakenly claims to be present in the original environment.

In the third stage, one judges that in the second environment, the individual could not have thoughts about arthritis, aluminum, sofas, as such. For we have imagined that the individual's own background knowledge does not discriminate arthritis, aluminum, and sofas from their respective substitutes. No one else has isolated arthritis in the imagined environment. No aluminum or sofas occur in the imagined environment for the individual to perceive. There is no one on whom the individual relies who has perceived these entities or who can specify them. One judges that under these circumstances an individual could not have thoughts such as that arthritis is a painful disease, aluminum is a light metal, or sofas are not religious icons.

The conclusion is that what thoughts an individual can have—indeed the nature of the individual's thoughts—depends partly on relations that the individual bears to the relevant environments. For we can imagine the individual's make-up invariant between the actual and counterfactual situations in all other ways pertinent to his psychology. What explains the possibility of thinking the thoughts in the first environment and the impossibility of thinking them in the second is a network of relations that the individual bears to his physical or social surroundings.

We started with recognition that certain cases are possible. One support for this recognition is memory of actual cases that are appropriately similar. The step carries the empirically warranted assumption that arthritis, aluminum, and so on, exist. One natural inference in the first step is from actuality to possibility. Here the first step has an empirical element.[26]

The second step is usually presented as also aimed at establishing a possibility. I think that sometimes possibility is at issue, but I think that it is never the key point of these thought experiments. I think that the key point in the second step is

[26] Is there further insight into possibility that is independent of an empirical assumption that there are relevantly similar actual cases? I do not know. But I will not pursue this question here. The relevant actuality need not be a particular case involving arthritis or aluminum. Although I assume that *arthritis* and *aluminum* are real kinds, the force of the argument does not depend on this assumption's being right. The concepts arthritis and aluminum are epistemic stand-ins for a wide range of successfully referential concepts. The argument works on the assumption that there are relevantly similar concepts that are representationally successful. The reflective power of the thought experiments does not depend on the empirical issue of exactly which concepts apply to real properties or kinds, although I do assume in the thought experiments that arthritis and aluminum are real. That power depends only on the idea that there are certain types of concepts and that a range of them are successful in representing properties or kinds. Thus the reflective judgment is that successful representational exercise of relevantly similar representational abilities cannot be as completely free of environmental connection—as would have to be the case if the same abilities were exercised in the second environmental scenario. One judges that different (equally successful) abilities are being exercised in that scenario. In the background, I think, is the principle that error presupposes a base of veridicality. I think that ultimately this principle is apriori. It may be that in addition to relatively obvious empirical warrants that we have for the first step, there are guiding apriori warrants that apply to whole classes of cases.

actually quite subtle. What is fundamentally at issue, in the second stage, is constitutive explanation, not modality. I shall return to this point.

The impossibility judgment in the third step derives from a fallible understanding of the types of conditions under which the relevant propositional abilities can arise. Specific propositional abilities are necessarily embedded in a network of other capacities and causal relations, evoked by examples in which they are stipulated to be absent.

The thought experiment concludes with a judgment about explaining differences between thoughts in the two environments. The second and third steps evoke a realization that physical differences in the individuals' bodies are not in general explanatory of the differences in thoughts. Putnam stipulates that the bodies are the same even though no water occurs in the counterfactual situation. This is, of course, impossible. But the main point of the step holds. Possibility is not what is fundamentally at issue. The point is to remind one that the nature of an individual's thoughts cannot be explained in terms of constituents of his cells or in terms of minor gravitational effects on his body. Even though, to imagine an individual in a waterless situation, one must imagine no water in the individual's cytoplasm and different gravitational effects on his body, these differences are not themselves relevant to explaining differences in the individual's psychology. Perceptual, communicative, cognitive differences—and macro-causal relations involved in these activities—are relevant resources for explanation. The insight is that whatever differences there may be in the thinker's bodies, these differences will not *in general* be the kind of differences that are relevant to explaining constitutive differences in thoughts. Reflection indicates that phenomenal and behavioral differences (understood in terms of physical movement) are also not key sources of explanation. The differences in thoughts constitutively depend on differences in macro interactions with aspects of relevant environments.

V

I want now to reflect on the kinds of insights involved in the philosophical thought experiments. In the meta-logic case, I cited four types of insights that illuminate inference. I will compare them with insights in the two philosophical cases.

All three cases involve reflective judgment about *the function of representational activity*. In the meta-logic case, the judgment is that the function of good deductive inference is to preserve truth. In the error-presupposes-veridicality case and anti-individualism cases, the judgment is that representational success in perception and belief is veridical specification of a subject matter. Cases in which the individual thinks of aluminum, where there is aluminum in the environment to be thought about, are cases of representational success. Such success fulfills representational function. The functional character of the matters to be explained

implicitly guides the eventual judgments about the type of individuative explanations that are necessary.

In the meta-logic and anti-individualism cases, insight into representational function is coupled with insight into generic *ways of fulfilling that function*. We know how to engage in good deductive inference. We know that specifying arthritis or aluminum depends on capacities to theorize about it, perceive it, or communicate with others who do these things.

The meta-logic and anti-individualism cases rely on insights into *the form of a representational activity and into relations among representational abilities*. Thus in the meta-logic case, there is insight into logical form, on which good deductive inference depends. In the anti-individualism case, one isolates from a thought arthritis is a painful disease the part that specifically concerns arthritis. Similarly, whereas the meta-logic case develops specific relations among inferential abilities underlying specific logical forms, the anti-individualism case depends on reflection on relations among cognitive capacities—perception, theory, interlocution.

The error-presupposes-veridicality case differs from the other two in these respects only because the principle being arrived at is so generic that insight into representational form, and into relations among representational abilities, is unnecessary.

All three cases essentially involve elementary insight into *necessary implications regarding the relevant subject matter*. The meta-logical case depends on recognizing relations between the concept of truth and relational semantical concepts like true of (or between the phenomenon of truth and the phenomenon of truth of). The philosophical cases depend on recognizing relations between belief and truth (and in the anti-individualism case, between concepts like arthritis and entities like arthritis).

Finally, all three cases involve insight into *explanatory priority*. The meta-logic case relies on recognizing that preservation of truth is to be explained in terms of semantical relations between logical form and subject-matter structure. The philosophical cases utilize, in different ways, recognition that having propositional attitudes is to be explained in terms of conditions in which such attitudes are successful. So, in the anti-individualism case, one recognizes that having certain thoughts is to be explained in terms of the successful cases in the first environment, and ultimately in terms of the exercise of cognitive capacities and macro-causal relations, underlying such exercise, to entities in that environment. We recognize on reflection that successful exercise of relevantly similar representational abilities cannot be free of environmental connection—as would have to be the case if the same abilities were exercised in the second-stage, counterfactual environment.

All these types of insight are important, and putting them together to yield new principles requires ingenuity and judgment. But the fourth type, insight into explanatory priority, seems to me to be the most striking one. The nature of

this insight is somewhat different as between the meta-logic case, on one hand, and the two philosophical cases, on the other.

In the meta-logic case, the idea is that insofar as logical consequence constitutively involves preservation of truth, and insofar as truth necessarily implies and is best constitutively explained in terms of truth-of and relations to a subject matter, logical consequence is to be constitutively explained partly in terms of relational *semantical* notions like true-of. The explanation is ultimately in terms of systematic assignments of entities from the subject matter to formal components of thought contents or sentences.

In the two philosophical cases, these points about truth are assumed as background, since mental states are type-identified partly in terms of their veridicality conditions. But the subject matter is not semantics but mental states. The explanatory relation is between subject matter and mental states. In both cases, the relation involves the primacy of veridical states. In the anti-individualism case, the explanatory claim invokes non-representational relations as well—causal relations in the empirical cases, constitutive relations in the logical cases.

The thought experiments exploit an insight into the nature of representational ability. Representational contents of attitudes help type-identify the attitudes. An ability to represent successfully, veridically, must be associated with conditions that enable that ability to be realized. An ability whose nature is to represent a specific property or kind must be specific to that property or kind—not equally appropriate to other properties or kinds. So the ability must be associated with *conditions* that are specific to that property or kind. The twin earth scenarios exploit this point. They show that there must be actual veridical relations, supplemented by either causal or constitutive relations, between mental states and subject matter if veridical (and hence non-veridical) mental states are to be explained and individuated. I will conclude by reflecting on what unifies these four types of reflective insight.

VI

A principle backing the classical view of reflection is that in having reason, a reasoner must be implicitly aware of norms for reasoning, if the reasoner is to know what he is doing; and to be reasonable, a reasoner must know what he is doing. The principle is supplemented by a metaphor: that reason responsibly sets its own norms and hence is able to reflect on what they are.

The metaphor covers an oversight. Reason is larger than reasoners. *Ideally*, or eventually, reasoners may set norms and reflect on what the norms are. But a given reasoner may not be able to go so far. I have cited two ways in which this shortfall may occur. First, animals and young children may lack the concepts of reason or of propositional attitude necessary to have the relevant insights. Second, even mature reasoners may not have conceptualized or otherwise unified

some norms for reasoning, even implicitly. Inaccessibility occurs in aspects of both apriori and empirical reasoning. So the traditional generalizations about why reflection is warranted are flawed. Reason is present phylogentically, ontogenetically, and historically before its full shape can be articulated by reasoners. Still, it seems to me that there are elements of truth in the classical view.

Let us reflect on critical reasoners. A *critical reasoner* is one who can suspend and correct propositional attitudes, conceived *as* attitudes and *as* being corrected, on the basis of reasons, acknowledged *as* reasons. A critical reasoner thus has concepts of truth, falsity, propositional attitude, and reason. By focusing on critical reasoners, I lay aside phylogenetic and ontogenetic difficulties: animals and young children are not capable of reflection. There remains the historical limitation. Resources necessary for successful reflection may require new knowledge or even the development of new concepts. Sometimes an individual develops new knowledge or concepts through reflection. Otherwise, fully successful reflection must await new concepts or knowledge developed by others—sometimes through reflection, often through other means.

We are critical reasoners. We are not in general guaranteed access to norms governing our rational activity. Still, in having the concepts of reason and propositional attitudes, a critical reasoner cannot be fully blind to the nature of his rational activity. Although a critical reasoner need not have a fully formed (implicit or explicit) conception of principles governing any of his critical reasoning, he or she must have a rudimentary grip on norms governing critical reasoning, a grip that involves at least partial conceptualization and some substantial conceptual know-how with the concept reason. If propositional attitudes, reasons, and the applications of reasons are to be conceptualized as such, there must be a capacity to recognize correct applications of reasons in *cases*. This gives a basis for inferring to patterns that constitute norms and principles governing those cases.

More is constitutively involved in being a critical reasoner. To have the concepts of reason, propositional attitude, and truth, many of the largest features of rational activity must be conceptually available to reflection. To get started in thinking about these matters, one must get certain things right. In the cases of these concepts, minimum competence requires a capacity to recognize on reflection certain basic necessary implications associated with these notions, or the phenomena that they apply to. The activity involved in critical reasoning requires having some conception of the point or function of the activity, the states and events that are constituent elements in this activity, and the main norms on and means of carrying out the activity. More concretely, critical reason requires having conceptions of the aims of belief and inference; the form and character of some inference rules and some propositional attitudes; the relations between belief, warrant, and truth; and certain relations among propositional forms and capacities.

All the insights in the meta-logical case and the philosophical cases that I discussed fit in this picture. All the cases are warranted by understanding.[27] I believe that some of them have a further source of warrant. Some are warranted inasmuch as they are necessary elements in the paradigmatically justified activity—critical reasoning. The recognition of the conceptual relations, the relations among psychological states and functions, and the insights into explanatory priority are warranted partly by being constituent elements in the whole critical reasoning process. They are justified inasmuch as they are necessary conditions of the possibility of critical rationality.[28]

A critical reasoner must have some capacity to understand the most fundamental functions and ends of critical reasoning, and must have at least a partial grip on the most general norms governing fulfillment of these functions and ends. To understand these norms, there must be some insight into the conditions under which they can be fulfilled. Insight into shoulds must be accompanied by some insight into cans.

Classical rationalism overestimated how specific and how far-reaching reflective apriori insight can be. I think that so far we have not learned just how much of the structure of critical reasoning is conceptually available to reflection. We have to find the limits of apriori reflection retrospectively—by looking at cases of successful reflection—rather than by trying to determine the scope and limits of such reflection in advance. Much apriori reflection is the result of induction, or genuine discovery, and of the development of new concepts. We can learn what reflection is successful, and what *sorts* of reflection are reliable, mainly by considering cases.[29]

A great deal of metaphysics has, I think, shown itself to be at best a merely speculative adjunct to the empirical or mathematical sciences. Armchair reasoning on the finity or infinity of the world, the nature of matter, problems about impact and continuous bodies, the nature of space and time, parts and wholes, the problems of abstract entities, has been and remain parts of metaphysics. But these enterprises can no longer be reasonably seen as in themselves sources of knowledge, except insofar as they unify what is known from the sciences, including mathematics.[30] They are not areas in which common sense or autonomous apriori intuitive reasoning by itself leads very far.

[27] The particular sub-types of understanding in the different cases, and corresponding particular sub-types of warrant, invite further investigation. At a sufficiently high level of abstraction, all seem to me to be apriori. But one cannot generate conclusions about what specific kinds of things exist from the thought experiments, together with knowing that one has concepts for specific kinds (kinds like *aluminum* and *arthritis*).

[28] This is not to say that there will not be additional sources of warrant for component parts. Since critical reasoning is the supreme authority for reason, its constituent operations obtain authority partly by virtue of being part of it in a way that the Supreme Court's decisions would have authority partly by virtue of its legal supremacy. If only it were an intrinsically rational enterprise!

[29] Obviously, there are many other types of apriori reflection besides those described here.

[30] This is broadly Kant's negative point in *Critique of Pure Reason*. I think that Kant's specific reasons for his view are not correct. But I believe that he was intuitively judging the consequences of

I believe that the meta-cognitive enterprise of reflection is different. Reflection can serve as an adjunct to any enterprise. But I believe it also offers substantive insights that are not parasitic on or merely supplemental to the natural or mathematical sciences. The insights are limited and fallible. Their nature remains to be better understood. But it seems to me that apriori reflection can yield limited autonomous insight, and even knowledge, in certain parts of semantics, philosophy of mind, and epistemology. This is because we are reasoning about reasoning itself. Apriori understanding in these areas is constitutive of understanding fundamental aspects of critical reason, and of us as critical reasoners.

Newtonian physics and mathematics for the methods of traditional metaphysics in areas where the subjects directly overlap. He was also right to believe that there was scope for more reflective methods in epistemology and in certain aspects of the theory of mind. I believe that here too his grounds for this belief were largely mistaken, but his instincts about areas in which apriori reflective knowledge is possible, or at least most likely to be fruitful, were broadly correct.