1 *Introduction*

In the essays collected here, I reflect on some types of cognition that rest primarily or entirely on understanding of thought or language.[1] Most of the understanding-based routes to cognition that I discuss are supported by non-empirical epistemic warrants. I focus on types of understanding, and types of non-empirical epistemic warrants, associated with self-knowledge, interlocution, reasoning, and reflection.

Three considerations motivate my interest in epistemic warrants associated with these types of cognition. First, the warrants that I discuss differ in interesting ways from epistemic warrants commonly scrutinized in philosophy—warrants for perceptual beliefs and for beliefs inferred from perceptual beliefs, and warrants for logical and mathematical beliefs. Second, I believe that the four types of cognition are, either in themselves or in significant respects, distinctive of persons, and (on earth) of human beings. Third, the warrants shed light on cognitive powers that underlie each type of cognition. The relevant cognitive powers involve a fascinating variety of kinds of understanding.

A *warrant* for a belief or other type of cognition is a right to rely on the cognition as cognition. A warrant derives from meeting standards for operating well in the service of fulfilling representational functions of the relevant type of cognition. The standards allow for limitation in available information and limitation in the individual's representational powers. Thus, even though being true is the fundamental representational function of belief, a belief can be warranted even if it is false—assuming that the belief is formed through a cognitive power that is reliably conducive to yielding true beliefs, given available information. So being warranted is fulfilling a less fundamental, but still important representational function—roughly the function of representing reliably and as well as is psychologically possible, given available information. Understanding a warrant

---

[1] I intend '*cognition*' as a somewhat open-ended term for levels of representation above the level of perception that are informative about the world. Above the level, in regards to sophistication and abstractness. Propositional attitudes are paradigm cases of cognition. I count lower levels of representation—such as allocentric, amodal spatial maps in memory—as cognition as well. Here I focus on propositional capacities. Since I take practical representation to have a cognitive dimension, I believe that there is both theoretical cognition about the physical, psychological, or mathematical world, and practical cognition about what is practically functional or good. Here I focus on "theoretical" cognition, commonsensical or scientific—that is, cognition whose primary representational function is to yield veridical representation of the world.

thus helps in understanding basic facts about the cognitive power and the power's representational functions. Understanding warrants can also help in understanding the natures of individuals with those powers. In the cases that I discuss, the main powers are powers of understanding. So the unifying project of these essays is to understand certain powers of understanding associated with the four types of cognition.

The first, second, and fourth types of cognition that I discuss (self-knowledge, interlocution, and reflection) are among the important ways in which human beings are psychologically distinct from other terrestrial animals.[2]

The third type (propositional reasoning) is probably shared with other animals. However, I think that it yields non-empirical, or apriori, cognition only in humans. I believe that some non-human terrestrial animals can reason. They depend for some of their reasoning on deductive inferential transitions that are warranted non-empirically. But I think it likely that all the starting points and endpoints of their reasoning are *empirical*. Their cognition is too closely connected to perception, and to action that serves their sensed needs, to rely on resources that are warranted purely by intellection.

Colloquially, a type of cognition is *non-empirical*, or *apriori*, if one's warrant for the cognition—one's right to rely on it as cognition—does not derive from sense experience. Apriori types of cognition rest on reason or understanding, that is, on intellection.

I am interested in non-empirical, or apriori*,* warrants for cognition because I believe that they are interesting in themselves and because I believe that they yield special insight into powers of cognition that are distinctive of human beings, or more generally persons. Most of the non-empirical warrants that I discuss with respect to all four types of cognition are, very likely, special to human beings. I think that these types of cognition are so central and fundamental that they help determine the psychological natures of individuals that have them. They certainly help determine the natures of selves and persons. Having non-empirical or apriori warrants depends on using powers of intellection that mark the distinctiveness of persons.

Leibniz thought that apriori cognition of necessary truths is the distinctive mark of human cognition.[3] It may well be one such mark. But there are others. Most of the cognition that I discuss here is not cognition of necessary truths. Much of it is nonetheless apriori. The essays seek to broaden awareness of the

---

[2] There may be an near ancestor of self-knowledge in higher animals—knowledge that they are in pain, for example. I think that such animals do not have a full-blown first-person I-concept, a concept of self. So the knowledge is not, strictly, self-knowledge. But they have some lower-level ego-centric analog of the first-person I-concept. I believe that conceptualization of such sensations as pain may well be a very primitive matter. I doubt that any such reflexive, knowledgeable, self-attribution is non-empirical. At any rate, the types of non-empirical self-knowledge that I center upon seem very probably not to occur in non-human terrestrial animals.

[3] See, for example, G. W. Leibniz, *New Essays on Human Understanding*, ed. and trans. P. Remnant and J. Bennett (New York: Cambridge University Press, 1981), I, i, 25.

range of cognition that rests on intellection, beyond the preserve of logic, mathematics, and other disciplines that bear on necessity.

I turn from colloquial specification of apriority to a more nearly precise specification. A type of cognition is apriori if and only if some instances of that type have non-empirical warrants. An apriori (non-empirical) warrant is one whose warranting force does not derive from perception or from sensing. Typically, the warranting force of an apriori warrant derives from intellection— reason or understanding. A psychological state, or a psychological transition, is apriori if the individual in that state, or making that transition, has and relies on an apriori warrant for it.[4]

Warrants derive from meeting norms, or standards, for operating well in the service of realizing a function or purpose. *Epistemic warrants* derive from a psychology's meeting norms or standards that govern good routes for realizing the representational function of belief formation—the function of forming true beliefs—or for realizing the representational function—preservation of truth and warrant—of certain transitions (inferences) that serve true belief. The relevant standard is for operating *representationally well* cognitively, well enough to have the right to hold the belief, or make the transition, given relevant information and relevant cognitive resources.[5] An epistemic warrant is a right to hold a belief, or to make an inference that serves true belief. Meeting the relevant norm, or standard, yields a right to hold belief or to make an inference. To *have* an epistemic warrant is to meet an epistemic norm or standard in a way that suffices for its being right to hold a belief or make an inference.

There are two types of warrant—justification and entitlement. A *justification* is a warrant that consists partly in the operation or possession of a reason. An individual is justified if and only if the reason is operative or relied upon in the individual's psychology.[6] An *entitlement* is a warrant whose force does not

---

[4] I understand apriori *truths* to be truths that can be believed with apriori warrant. It may be that any apriori truth can also be believed with empirical warrant. But not all truths believed with empirical warrant can be believed with an apriori warrant. I think that warrant for a cognitive state instance, or (even more fundamentally) for a cognitive state type relative to a way of having or forming that cognitive state type—together with warrants for inferential transitions—are the basic sorts subjects of apriority. Apriori warrants for ways of having psychological states or ways of making inferential transitions are basic. Apriori truth is to be explicated in terms of apriori warrant. I believe that traditional rationalists, who often wrote of apriori truths, conceived such truths as apriori because those rationalists took such truths to be associated with canonical, warranted ways of establishing them—through a proof, for example. For further discussion of these matters, see my 'Frege on Apriority' in *Truth, Thought, Reason: Essays on Frege* (Oxford: Clarendon Press, 2005), 357–364.

[5] A cognition's operating well representationally is to be distinguished from its operating well practically. Its operating well representationally is serving the function of forming, or preserving, true belief. Its operating well practically is serving the function of forming states that benefit the individual's survival for mating, or that serve some other practical—perhaps moral—function or purpose. See my *Origins of Objectivity* (Oxford: Oxford University Press, 2010), chapter 8.

[6] The simplest type of availability is presence and accessibility within the prospective believer's own psychology. But individuals can have justifications even though they lack the justifying reason as the content of a fully formed state in their psychologies–for example, if they could through prompting or reflection come to form such states relatively easily. I discuss such matters in 'Some Origins of Self'

consist, even partly, in the individual's using or having a reason. Entitlements can be rationalized—associated with a justifying argument that explains the entitlement. But entitlements do not require the individual to have such reasons or explanations. An individual has an entitlement if and only if the entitlement attaches to the psychological processes that yield the relevant state or transition. Entitlements need not be understood, understandable, or even thinkable, by the warranted individual. They need not reside in any reasoning that is operative or relied upon in the warranted individual. They derive purely from the individual's psychological system's operating well with regard to realizing or fulfilling its representational functions.

Each type comes in epistemic and practical forms. I shall focus on epistemic forms. *Epistemic* entitlements and justifications derive from operating well in the service of realizing the representational function of forming true beliefs, or carrying out inferences that function to support the truth of beliefs.

Many epistemic warrants suffice for *knowledge*, if they are undefeated by stronger warrants and if there are no other countervailing circumstances.

Entitlements as well as justifications can be apriori. An entitlement is apriori if the correct account of the source of the entitlement adverts entirely to powers of understanding or reason, even though the entitlement does not depend on the warranted individual's having the understanding or reason. For example, the account may say that an individual is entitled to rely on an inference in reasoning, because the inference rule that in fact governs the inference is apriori truth- and warrant-preserving, even though the individual could not think the inference rule, and thus is unable to use it as a reason. Or an individual may be apriori entitled to rely on what he or she is told as being true because the form of the telling is an apriori sign of rationality in the source, even though the individual recipient does not and could not have this account as a reason for relying on what he or she is told.

I emphasize two traditional points about apriority. I do so because they are still often missed.

First, to be apriori is not *eo ipso* to be infallible, indubitable, or unrevisable. An apriori warrant is one that does not derive its force from sense experience. It does not follow that the warrant *infallibly* guarantees that a warranted belief is true. One can have an apriori warrant—short of a proof—to believe a complex mathematical proposition; but the belief may be mistaken.

For similar reasons, one can have an apriori warrant, but *doubt* the proposition that one is warranted in holding. So having the warrant for believing a proposition

---

and 'Self and Constitutive Norms', reprinted later in Part I. There may be cases in which fully formed psychological states operate as reasons in a psychology, but are not accessible, even in principle, to the reasoner's consciousness. That is, there may be modular reasons. In any such cases, the reasoner is, of course, not responsible for the modular reasons or reasoning. I think of such cases of warrant as nearly at the borderline as between justification and entitlement. I see them, however, as justifications within the individual's psychology even if they are not justifications available to the individual. If one wanted, one could say that the *individual* is entitled to the justification that occurs within the inaccessible parts of the individual's psychology.

does not entail that the proposition is indubitable. One may mistakenly not rely on one's warrant, either because one is psychologically timid, or because one incompletely understands the proposition that one is warranted in believing, or because one has mistaken views about the warrant itself. Or one may find countervailing considerations that defeat the considerations that yielded a warranted belief. Then one might *rationally* doubt the proposition that was once apriori warranted.

Because some apriori warrants do not guarantee the truth of propositions that they warrant belief in, they are vulnerable to possible counter-considerations. That is why some beliefs that have apriori warrant are rationally *revisable*. I think that apriori warranted beliefs can be revised even on *empirical* grounds. Cases of this sort arise in certain types of reflection.

The main point to bear in mind here is that a claim that a type of cognition, or an epistemic warrant, is apriori is a claim purely about the nature of the standard that is met in having a right to hold a belief. Although claims of apriority in philosophy have often been associated with infallibility, indubitability, or unrevisability— indeed, with close-mindedness—apriority *per se* has no such associations.

I do believe that a *few* apriori warrants are infallible and cannot be rationally revised. They cannot be defeated by any other warranted consideration. I think that our warrants to believe instances of pure *cogito* thoughts and some warrants to believe certain simple logical and mathematical truths, or to rely on certain simple forms of inference, have these features. Most of the apriori warrants that I discuss are not even close to infallibly guaranteeing truth, or to being rationally unrevisable.

I doubt that *any* proposition is strictly indubitable. Psychologies are very malleable. In any case, apriority concerns the nature of one's right to believe, not one's power to resist belief.

There may be propositions that are not *rationally* dubitable. But these are not co-extensive with apriori knowable propositions. One can be apriori warranted— for example, in believing a complex, inductively supported mathematical conjecture—and rationally doubt whether one has gotten things right.

The second traditional point about apriority is that one does not show a warrant not to be apriori by showing that the warranted belief or thought-transition depends on sense experience. A warrant is apriori if the force of its warrant is independent of sense experience. That sort of independence is compatible with various other ways in which a warranted belief or thought-transition may depend on sense experience, besides depending on sense experience for the force or nature of the relevant warrant. For example, to acquire a given belief or to understand the believed proposition, one might have to read a book, study a diagram, or hear an explanation. The relevant sort of independence of sense experience concerns the force or nature of the warrant, not whether having the warrant, or the belief, depends causally on having sense experience. The warrant itself must not advert to sense experience, even if having the warrant psychologically depends on having had sense experience.

An apriori warranted belief can also be empirically warranted. So even showing that a belief has a warrant whose force derives from sense experience does not show that the belief is not apriori warranted. Apriority ultimately concerns the nature of a specific warrant, not the nature of a truth, or a belief. We have apriori warrant to believe simple arithmetical truths, through understanding them. There may be a secondary empirical warrant for such beliefs—from counting, or from the role of arithmetic in empirical science—as well.

The focus in most of these essays is epistemological. My main motivations lie, however, in a desire to understand certain psychologically relevant differences between persons, including human beings, and other animals. As I noted in the Preface, I think that the apriori warranted human cognition that I discuss here is not shared with other animals. Fulfilling epistemic norms—and having epistemic, including apriori, warrants—depends on good use of relevant cognitive powers. By reflecting on good, warranted use to the underlying powers, one can learn something about the natures and representational functions of the powers. Thus the essays point toward understanding distinctively human psychological capacities. More generally, they point toward understanding psychological capacities distinctive of persons.

Of course, there are many distinctively human, cognitive—or cognitively dependent—capacities beyond those that I discuss here. I barely touch on human knowledge of language, mathematics, and history. I have almost nothing to say about human capacities to create art, to form societies governed by legal norms, or to maintain personal relationships that involve understanding others' points of view. Most of these capacities employ one or another distinctively human type of cognition. I hope to say more about some of these matters eventually. I do not, however, propose, now or later, to survey all ways in which human beings are psychologically or socially distinctive.

Humans share perception with a vast range of other animals. I think it likely that humans also share perceptual belief and simple propositional inferences involving such beliefs. In other work, I explore these shared types of representation and cognition.[7]

Here, I center on four types of cognition that are non-empirical, or have substantial non-empirical aspects. Each of these types also has uses and applications that are empirical. We know a lot about ourselves, for example, by making inductions from self-observation. Much of our non-inferential introspective self-knowledge is empirically grounded.

Similarly, the vast bulk of what we know from being told things by others utilizes empirical warrants for trusting others' expertise, or utilizes empirical background knowledge to separate questionable from trustworthy elements in others' reporting.

---

[7] For background on shared types of cognition, see my *Origins of Objectivity*; 'Perceptual Entitlement', *Philosophy and Phenomenological Research* 67 (2003), 503–548; 'Origins of Perception', *Disputatio* 4:29 (2011), 1–38; 'Steps Toward Origins of Propositional Thought', *Disputatio* 4:29 (2011), 39–67.

I now believe that *all* interlocution is warranted empirically. I believe that our warranted comprehension of what others say or write inevitably has an empirical element. (I discuss this change of view below and in 'Postscript: Content Preservation'.) I continue to think, however, that there are substantial and interesting non-empirical components to warrants for relying on what others tell one.

Inductive reasoning commonly has empirical warrant, even apart from any empirical premises used in the reasoning. That is, many—indeed most—inductive inferential transitions are warranted empirically. Clearly, however, some reasoning rests on inferential transitions that are apriori warranted—for example, deductive reasoning or some inductive, inference-to-the-best-explanation reasoning in mathematics.

Much—probably most—successful reflection, including philosophical reflection, has broad empirical warrant. Still, some reflection depends on cognitive states with apriori warrant.

So, in considering non-empirical cognition in these four areas, one must isolate lines of warrant that are non-empirical from other lines that are empirical.

Some philosophers are empiricists. They believe that there is no significant knowledge or belief that is warranted non-empirically. Empiricism dominated philosophy in the Anglo-American tradition from the 1920s into the latter part of the twentieth century. But empiricism never gave a plausible account of logical, mathematical, or moral knowledge—each constituting a large area of human cognitive life. This fact came to be recognized increasingly in the last two decades of the twentieth century.

Empiricism is no longer a dominant doctrine. I do not focus on criticizing it. In these essays, I *assume* that empiricism is mistaken. Its inability to account for logical, mathematical, and moral knowledge, and certain *cogito* types of self-knowledge, seems to me clear. I explore types of non-empirical cognition that go beyond the most obvious types. Since some of the understanding that I discuss *is* empirical (in particular, comprehension of language utterances by others), it is understanding, not apriority, that provides the unifying theme in the four areas indicated in the volume's title. Still, the main focus of these essays is on types of understanding that are warranted non-empirically.

I

Part I of this volume centers on self-knowledge.[8] As noted, a lot of our self-knowledge is empirical. Descartes' *cogito* (I am now thinking) exemplifies, however, a type of self-knowledge that is non-empirical. Descartes' famous

---

[8] For a source of writing on self-knowledge that I do not reprint in these essays, see M. J. Frapolli and E. Romero (eds.), *Meaning, Basic Self-Knowledge, and Mind: Essays on Tyler Burge*, (Stanford: CSLI Publications, 2003). The volume contains my replies to eleven essays on my work, many of which center on issues regarding self-knowledge. I do reprint one reply that concerns reflection. See Part IV in this volume.

example suggests a fairly wide range of further types of non-empirical self-knowledge. I believe that these types help mark human beings—especially their moral powers and powers for critical reasoning—as distinctive. The essays in Part I gradually develop a unified account of non-empirical types of self-knowledge.

'Individualism and Self-Knowledge' (1988) was written with a single objective in mind. It was to show that anti-individualism and non-empirical types of self-knowledge are compatible. Anti-individualism is the view that the natures of many psychological or mental states are partly determined by relations between individuals and specific attributes in the environment.[9] Non-empirical types of self-knowledge hinge on warrants that support knowledge, but that do not rely on sense experience for such support. For the most part, we can know what the environment is only through empirical warrants—warrants that depend for their force ultimately on warranted perceptual belief and on warranted inferences from such belief. So prima facie, it can seem puzzling that we can know non-empirically what our mental states are—which involves knowing their representational natures, their representational contents—when those natures depend for being what they are on relations to specific attributes in the environment that can be known about only empirically.

The basic idea for resolving the puzzle is that self-knowledge, no less than ordinary knowledge of the environment, depends on utilizing a standpoint whose representational resources have already been made into what they are by the environment. In knowing the representational contents of our mental states we commonly use those very contents in specifying themselves and in specifying the mental states that they help type-individuate. We know how to use and specify those contents by understanding them, not by investigating their constitutive conditions. In other words, at the point where we use reflection for the relevant sort of self-knowledge, the environment has inevitably already had its effect in making the psychological reflection (and the psychological states which are reflected upon) what they are.

To understand those contents in a way sufficient to know what they are, we need not distinguish them from alternatives that might have been our contents if we had come to have our psychological states and contents in a different environment. So although we can know only empirically some or all of the features of our environment that determine the natures (including representational contents) of our psychological states, we can know what our psychological states (and their contents) are non-empirically. The essay brings out the fact that we have no standpoint to know ourselves beyond the standpoint that is marked by being embedded in the environment that we find ourselves.

---

[9] For fuller discussion, see the Introduction to my *Foundations of Mind Philosophical Essays, Volume II* (New York: Oxford University Press, 2007), and several essays in that volume, and *Origins of Objectivity*, chapter 3.

In writing the essay, I categorized certain types of non-empirical self-knowledge, radiating out from Descartes' pure *cogito*: I am hereby engaging in thinking. I stated that I thought that these pure cases contain seeds for developing an understanding of other non-empirical types of self-knowledge. Many readers—understandably interested in the wider topic of self-knowledge, non-empirical and otherwise—leapt to the conclusion that I was attempting to give a general account of self-knowledge, or at least of non-empirical types of self-knowledge. This view led to many observations of at least prima facie difficulties in taking what I said in the essay to provide such an account. Many of these observations are correct. In fact, I articulate several of these difficulties in the essay. Presented as criticisms, they were irrelevant. For my objective in the essay was not to give a general account of non-empirical types of self-knowledge. It was to show why the non-empiricality of certain types of self-knowledge is compatible with anti-individualism.

It was often pointed out, in response to the article, that to know that one believes that cats are animals, one must not only be warranted in one's specification of the representational content of the belief (the content—cats are animals). One must also be warranted in specifying one's mental state as a belief, as distinguished from a hope or a wish. I said nothing about the warrant for this latter specification.

To take another example, my account centers on self-knowledge of states with representational content, particularly propositional attitudes. As I noted myself, it says little about knowledge of one's sensations, such as knowledge that one is in pain. I believe that such self-knowledge is empirical. Still, it shares many striking features with non-empirical self-knowledge—its directness, relative certainty, and so on.

I did not discuss such matters, except to note them. I did not try to fit them into a general account of self-knowledge. My aim was to show the compatibility of anti-individualism and non-empirical types of self-knowledge. The puzzle about such compatibility derives from the way the *representational contents* of one's first-order mental states depend on the environment. Determining the mode (belief, hope, wish) of one's psychological states raises its own questions. But it does not raise questions about compatibility with anti-individualism. The apparent threat to compatibility derives from the way in which representational content of mental states depends on the environment.

Similarly, I did not try to fit self-knowledge of sensations into a general account, because knowing what one's sensations are does not hinge on specification of representational content, and does not (at least in some key respects) depend for its nature on relations to an environment beyond the individual. I took accounting for the *general* phenomenon of authoritative and (as a sub-set) non-empirical types of self-knowledge to be a further project.

I did take myself to be setting a direction for developing a more general account. I thought—and still think—that by reflecting on Descartes' *cogito* case, and mining it for its deeper implications, one can gain insight into other

types of self-knowledge that do not share some of its particular features. *Cogito* is commonly taken to be a curiosity. It is infallible and self-verifying in ways that no other types of self-knowledge are. However, it either embodies or points to characteristics of the broader range of other non-empirical types of self-knowledge. I return to this point at the end of this section.

'Our Entitlement to Self-Knowledge' (1996) discusses another feature of non-empirically warranted self-knowledge. This feature is its symbiotic relation to critical reasoning. I believe that Descartes' reflexive *cogito* and other instances of non-empirical self-knowledge are necessarily associated with a structure of critical reasoning. Critical reasoning is the process of evaluating reasons as reasons, and implementing those evaluations in thought and action. The structure of critical reasoning consists in apriori connections—such as inferential transitions—among different types of states constitutively involved in critical reasoning. I believe that the structure type-individuates the type of thinking that constitutes critical reasoning, and the structure provides a type of unity that marks an individual as a critical reasoner.

The essay sets out a structure of transitions within critical reasoning. When thought follows procedures associated with this structure, the thought is non-empirically warranted. I claim that the structure is an integral part of, constitutive to, critical reasoning. Following procedures associated with the structure requires making transitions based on *understanding* reasons and their applicability to propositional attitudes. The relevant warrants derive partly from the minimal-competence understanding that is involved in engaging in critical reasoning.

Discussed in the abstract, the structure can sound distant and formidable. In fact, it is familiar. It is simply what is common to instances of reasoning like the following: I believe that *p*; but wait, I have obtained reason to believe that *r* and that *s*. I believe that *r* and that *s*. Taken together, *r* and *s* count as strong reasons against believing *p*. I know of no sufficient reasons that counteract or outweigh these reasons against believing *p*. So I have sufficient reason to believe not-*p*. So not-*p*.

What is distinctive about the structure is that it cites reasons *as* reasons for or against one's own attitudes, which one also reasons about; and following it involves implementing one's reasoning about reasons, for or against one's attitudes, into the attitudes one reasons about. One implements the second-order reasoning about first-order attitudes in the first-order attitudes (the last step). I believe that we know by apriori reflection that our reasoning sometimes has this structure. And I think that we know apriori that reasoning that follows this structure is (critically) reasonable, in this sense: If one is reasonable (and warranted) in accepting the individual steps in reasoning that has this structure, then one one is reasonable in one's conclusion.

The argument of the essay hinges on distinguishing the sort of *self-knowledge* that is a key element in critical reasoning from self-knowledge that is grounded in empirical self-observation. A key distinguishing feature is that the exercise of capacities underlying empirically grounded self-knowledge allows for *brute error*. Brute error is the type of error that is compatible with being warranted in

one's belief and that does not derive from any sort of psychological malfunction. Intuitively, brute error in self-attributions derives not from internal shortcomings, from normal inductive errors, or from misperceptions of behavior that derive from abnormal observation conditions. The type of self-attribution that underwrites the self-knowledge that is integral to critical reasoning does not allow *brute* error: all errors derive from some sort of misuse of the competence that enables self-knowledge to figure in the distinctive structure of critical reasoning. Such self-attribution is not empirically grounded; it does not rest on self-observation, much less inference from self-observation.

The idea is that if a type of self-attribution yields the relevant self-knowledge when the self-attribution is successful, and if that type is constitutively necessary for the rational review of one's attitudes in critical reasoning, then that type must be immune to brute error. Otherwise, there would be a dissociation between the rational review of one's attitudes and the implementation of the reasons that are marshaled in the review within the set of attitudes that are under review (the lower-level, first-order attitudes). The last step in the structure that I laid out earlier would not follow apriori from the preceding steps (assuming them warranted/reasonable). This point is worked out in some detail in the article. The argument emphasizes rational, non-empirical connections within the structure of critical reasoning. Given that it is apriori reasonable to believe in critical reasoning, with the structure of reason-yielding transitions that the argument lays out, the argument yields an apriori reason to believe that some self-knowledge is warranted non-empirically.

The argument of the essay was meant to support not only the view that to figure in critical reasoning in the way that some self-knowledge does, such self-knowledge must be warranted non-empirically. It was also meant to support the view that part of the non-empirical warrant for the relevant type of self-knowledge derives its force from the role of such self-knowledge within the rational structure of critical reasoning.

Christopher Peacocke accepted the idea that some self-knowledge has non-empirical warrant, and must have such warrant to figure in critical reasoning. He maintained, correctly, that this point is prima facie independent of the further point that the non-empirical warrant for the relevant self-knowledge derives its force from something about critical reasoning. Peacocke was not persuaded by the second point. He held that the relevant self-knowledge derives its non-empirical warrant from sources that are completely independent from the role that that self-knowledge has in critical reasoning.[10]

I thought that the competence understanding involved in the relevant self-knowledge gains some of its warrant from its functioning within the structure of critical reasoning. Indeed, I thought that the understanding draws on fundamental transitional capacities that are integral to being a critical reasoner. This point invited further development. Later essays engaged in such development,

---

[10] Christopher Peacocke, 'Our Entitlement to Self-Knowledge: Entitlement, Self-Knowledge, and Conceptual Re-Deployment', *Proceedings of the Aristotelian Society* 96 (1996), 117–158.

culminating in the third of the Dewey Lectures. I return to the issue at the end of this section.

'Our Entitlement to Self-Knowledge' connects closely with the Dewey Lectures and with essays in Part III of this collection. It is a starting point for 'Reason and the First-Person'.

'Memory and Self-Knowledge' criticizes errors in Paul Boghossian's attempt to show that anti-individualism and non-empirical self-knowledge are incompatible. Much of the discussion turns on a distinction, which I had made earlier in 'Content Preservation' (see Part II of these essays), between *substantive memory* and *purely preservative memory*. Purely preservative memory is a type of memory that functions to preserve past thoughts for current use, without introducing new contents or attitudes (for example, as premises), with their own warrants and subject matter, into current cognition. Substantive memory is memory that does introduce new premises, with their own warrants and subject matter, into current cognition.

I discuss several errors in Boghossian's argument. The key error is the unargued and question-begging assumption that lacking an ability to identificationally discriminate one past content from another, one cannot know one's yesterday's thought. But knowing one's yesterday's thought does not, in general, depend on identificational discrimination of one past content from another. It is not an identificational representational competence. It commonly depends only on competently preserving the past thought in memory. If purely preservative memory operates properly, it preserves knowledge automatically.

In addition to the unsound abstract argument that anti-individualism and non-empirical self-knowledge are incompatible, Boghossian presents a parallel argument that anti-individualism is incompatible with our assumption of an ability 'to tell apriori whether any particular inference of ours satisfies one of [the forms of valid inference]'. Boghossian's view rests on use of the slow-switching case that I introduced in 'Individualism and Self-Knowledge'.

Let us suppose, in accord with my case, that an individual might have "twin" concepts that the individual is not in a position to discern from one another, because the individual does not realize that he or she switched environments. The idea is that such an individual might carry out an inference that begins with one of the twin concepts in the first premise, introduces the other twin concept in the second premise, and applies an inference rule that requires a single concept. Such reasoning would yield an invalid argument, whose invalidity depends on the equivocation. Boghossian concludes that the possibility of such a case undermines our ability to tell apriori whether any of our arguments is valid.

This case raises more interesting issues than the first argument. It should immediately strike one, however, that the conclusion is overwrought. The mere possibility of such a highly unusual case hardly undermines our normal ability to determine non-empirically that our inferences are valid. At most it might show that such ability is in principle fallible, and that a default apriori warrant could in principle be overturned by empirical considerations (considerations that showed that the circumstances of the argument were abnormal). It seems to me that

relatively little of the work in the cottage industry that has developed out of discussion of these issues shows a perspective on the *outre* character of the cases that are discussed. The value of the cases lies in sharpening our conception of the powers of reason, not in adjudicating the truth of large doctrines such as anti-individualism, the apriority of reasoning, or the existence of non-empirical self-knowledge.

In 'Memory and Self-Knowledge', I note that there is always a route to avoiding equivocation. One can resolutely rely on purely preservative memory to connect later uses of a concept anaphorically to the first use in the argument. I certainly do not think that in every imagined case, an individual thus relies on purely preservative memory. I take the existence of purely preservative memory, and our ability to resolutely rely on it, to show that the slow switching cases cannot prevent a carefully directed reasoning process from yielding valid reasoning. There is no across-the-board undermining of the control of reason in processes of reasoning.

I think, however, that the relevant type of equivocation *can* occur. Individuals do not always rely on purely preservative memory. That point is obvious from the fact that in *ordinary* cases, reasoners sometimes fall into fallacies of equivocation. The switching cases elicit the fact that there can be a subtle interplay between the external circumstances that motivate use of a premise in an argument and the employment of purely preservative memory.

Again, this sort of interplay occurs in ordinary cases, particularly cases involving occurrence-based, context-dependent representational devices. For example, one can think that ball is red, looking at a red ball; then one can think, after blinking but again looking at a ball, that ball is round; then one concludes, that ball is red and round. Suppose that while one blinked, a different ball is substituted. The new ball is, in the context, indiscernible from the first ball. Suppose, however, that the new ball is not red, though it looks just as red as the original ball. Then the truth value of the conclusion and the validity of the argument hinge on what one's demonstrative thoughts in the second premise and conclusion refer to.

The issue illustrates tension between purely preservative memory—which would anaphorically retain reference between the first and second premises—and environmental pull, from perception, on reference—which would divert demonstrative reference in the second premise from the referent in the first premise to a different referent, the new ball that is perceived. If one resolutely attaches the second occurrence of the form that ball (the occurrence in the second premise) to the occurrence in the first premise, relying on the first occurrence to establish the referent in the second, then the argument is valid. In such a case, anaphora overrides perception in establishing the referent of that ball in the second premise. If, as is more common and natural, one relies on perception to establish the referent of each occurrence of the form that ball, then the argument is, as it stands, invalid; and the conclusion is false.

In such a case, one would be presuming that the referent of that ball in the second premise would be anaphorically connected to the occurrence of that ball in the first premise. Otherwise, one would not have taken oneself to be producing a deductive argument. Environmental pull, deriving from the context-dependent referential force of perception, would render one's presumption mistaken. One would then have engaged in an unnoticed switch in referents that would render one's argument invalid—despite one's logical competence and apparent reasonability in carrying out the argument.

On reflection, one should realize that *whenever* one relies on a sequence of demonstrative- or indexical referential devices in establishing a referent in a context, one is in principle liable to undiscerned reference shifts. This is a recurrent, normal source of possible invalidity in deductive argumentation. We commonly handle such cases by presuming on our normal, highly reliable capacities for contextual object tracking. Such capacities in effect presume identities of reference linking different occurrent applications of demonstrative or indexical representational contents in thought. Such identities are normally not consciously added as separate premises. Even when the identities fail, presuming upon them yields something like warranted, valid, but unsound reasoning. Cashing out this term "something like" is non-trivial. The reasoner is normally warranted, and reasons competently, even when unrecognized referent switches occur.

The switching cases introduced in 'Individualism and Self-Knowledge' are more radical than the switching case just discussed. Assuming that enough time passes after a switch, and assuming other specific conditions are met (such as that memories of the old environment are not lost), the changes in reference that are involved in switches of whole environments—not just contextual referents in a given environment—yield a different sort of change in reference. The change is not simply a change in the referents of demonstrative- or indexical referential devices.[11] The change is in ability-general representational contents, concepts.

In the most interesting switching cases, the individual has two concepts that are not introspectively discernible to the individual. In the relevant argumentation, the reasoner does not resolutely rely on purely preservative memory. One premise calls up one of the concepts; another premise calls up another. And deductive argument fails because of the switch.

The idea that an individual might have two concepts that the individual cannot discern by simple introspection is sometimes presented as if it were a peculiar, or even non-credible situation. However, concepts mark conceptual modes of presentation that mark psychological abilities. Many differences in psychological abilities are not open to immediate introspective discernment. There are many aspects of our psychological states that we cannot cognize easily and through

---

[11] See 'Other Bodies' in Andrew Woodfield (ed.), *Thought and Object: Essays on Intentionality* (Oxford: Clarendon Press, 1982), 97–120, reprinted in *Foundations of Mind* (Oxford: Clarendon Press, 2007), for extended discussion of this point.

simple reflection. The idea, often presented as common sense, that the nature of our thoughts is transparent to us is, I think, simply naive.

The cases in which concepts are switched are more unusual, and hypothetical, than the cases in which the referents of demonstrative or indexical representations are switched. Nevertheless, the same sort of tension between purely preservative memory and environmental pulls on reference occur. In the concept switching cases, purely preservative memory has a stronger pull than in the contextual-referent switching cases. But one can, I think, certainly imagine cases in which concept switches occur in the course of a deductive argument, undermining the argument.

In 'Memory and Self-Knowledge' I note that there is a significant difference between the "equivocations" that are imagined as a result of slow-switching and the equivocations that exemplify ordinary errors in inference. I stated that "equivocations" in the slow-switching cases are always cases in which the reasoner 'tacitly and mistakenly presupposes that the [twin] concepts apply to the same objects'. This tacit presupposition, or presumption, is shown by three facts. First, the reasoner does not realize that there are different concepts; the concepts apply to different entities that are indiscernible to the reasoner at the time of the reasoning. Second, the reasoner would accept the results of substituting either concept for the other in a generalization, or descriptive predication. Third, if the reasoner were in circumstances in which use of one of the "twin" concepts were be elicited by a question, and the reasoner were presented with an object that satisfies the other concept, the reasoner would (mistakenly) apply the concept to that object. In short, except for the important fact that the reasoner's psychology may never unite the two concepts in an equivalence claim, the reasoner's psychology shows every sign of being disposed to accept 'For all objects x, Fx if and only if F_tx', where the term that 'F_t' stands in for expresses the "twin" concept of the concept expressed by the term that 'F' stands in for.

Exactly how to characterize this "presupposition" or "presumption" seems to me an interesting psychological question. I do not try to give a precise answer to the question. What seems to me clear, however, is that the presumption is warranted, given the thinker's information. The thinker would be entitled to take instances of the twin concept (say, Tw-water) to be instances of the home concept (instances of water). The existence of the warranted "presumption" allows mistakes (the analogs of equivocation) in argumentation that derive from slow-switches to be assimilated more to empirical errors than to unreasonable inferences. The form of the inference would involve an ordinary equivocation, *if* one omitted to include the presumption of identity, or sameness in tracking. The presumption makes the inference more like a valid enthymeme with a mistaken tacit "premise" than like an ordinary fallacy of equivocation.

As in the case of demonstrative tracking, I do not claim that there *is* a mistaken tacit premise. The presupposition or presumption must be characterized more subtly. Still, the errors appear to me to be warranted. And it appears that the

mistakes are not failures in a specifically logical competence. They are more like warranted, unfortunate failures in tracking. They are warranted, empirically based failures that affect the application of one's logical competence.

Their possibility brings out in-principle limits to our non-empirical capacities to understand our own minds infallibly. They do not undermine the intuitive point that in normal cases we can determine non-empirically the validity of our arguments—even when we are *not* resolutely relying on purely preservative memory. In normal cases, we correctly understand the sequence of steps in our deductive arguments as involving univocality, or sameness, of concepts through the steps. Unless we have specific grounds to doubt that the cases are normal, we are apriori warranted in relying on our intellectual competence.

Determining the exact normative status of the equivalence presupposition or presumption is an interesting project.[12] The normative status of the reasoning in slow-switching, "equivocating" cases is quite general, not local to particular imaginable cases. The role of the equivalence presupposition, or presumption, shows that, apart from resolute reliance on purely preservative memory, we have no infallible, transparent competence to avoid such errors. Reasoners that incur the imagined "equivocations" are not thereby irrational. Intuitively, their inferences remain apriori warranted. Apriority consists in warranted, not infallible, reliance on understanding or reason. In normal cases, and even in most switching cases, competent reasoners are warranted in relying on their understanding in producing formally valid arguments.

'A Century of Deflation and a Moment about Self-Knowledge' was the last Presidential address to the Pacific Division of the American Philosophical Association in the twentieth century. In the first half of the address, I used the occasion to spoof deflationary tendencies in philosophy during the century.

Philosophy can no longer arrogate to itself a central role in yielding knowledge. The sciences have long separated themselves as autonomous disciplines and eclipsed it in driving cognitive progress. I think that philosophy still makes progress, and does sometimes yield knowledge, on its own. It makes progress in contributing to scientific advances—especially in frontier sciences—by clarifying concepts and isolating presuppositions in scientific knowledge. It yields knowledge of its own in areas of cognition that are not systematized by a science. Epistemology is one such area. For example, understanding at least some types of self-knowledge is something philosophy is perhaps uniquely fitted for. Moreover, philosophy can contribute clarity and depth of understanding—including understanding unclarities and difficulties—at levels of abstraction rarely confronted in the sciences. So it has a unique and valuable role, even though, because of the sheer complexity of what is known, its place in intellectual culture cannot be as central as it once was.

---

[12] My students Mikkel Gerken and Luca Struble have produced interesting and different work on this issue.

The analytical part of the essay takes up the issue raised by Peacocke that I discuss above in connection with 'Our Entitlement to Self-Knowledge'. I explain why I think that Peacocke's account of the special "authoritative" nature of non-empirical self-knowledge is inadequate. I claim that some of the special authority of non-empirical self-knowledge lies in another feature implicit in Descartes' *cogito*—mastery of a mature first-person concept. I maintain that 'the ability to move freely, rationally, and immediately, from first-level attitudes to second-level self-attributions of them, and back'—a crucial aspect of the structure of critical reason set out in 'Our Entitlement'—is part of the understanding involved in mastering the first-person concept. Such understanding is part of the source of warrant for non-empirical self-attribution.

There are two aspects to this source. One is the capacity to understand and adjudicate threats to one's prima facie warranted self-attributions. A person is an authoritative judge partly because persons "know the law"—at least some of the rational norms by which hard cases are adjudicated. The other aspect is that mastery of the first-person concept partly consists in applications of it that involve acknowledging intellectual and moral responsibility for certain instances of one's psychological states. This aspect is the genus of which the performative aspect in Descartes' *cogito* is a species. The element of being responsible for a thought within a system of critical reason is non-empirical and authoritative, other things equal.

Of course, neither of these aspects of a source for authority in non-empirical self-knowledge can make a self-attribution warranted unless the understanding being exercised is reliable in yielding veridical cognition. However, exercise of the relevant type of understanding within the structure of critical reason owes nothing for its warrant to sense experience.

I think that although the discussion in this essay is right as far as it goes, it does not sufficiently articulate wherein non-empirically warranted self-knowledge is immune to brute error, or wherein that sort of self-knowledge gets its warrant partly through its association with critical reason. There are elements of rational structures that inform self-knowledge of representational states that deserve more explication. I make glancing reference to these elements in the third-to-last paragraph of the essay. I do not elaborate them until the last of the Dewey Lectures, 'Self-Understanding'.

In 'Mental Agency in Authoritative Self-Knowledge: Reply to Kobes', I respond to Bernard Kobes's fine development of a connection between the performative element in *cogito*-like instances of self-knowledge (instances that I call 'basic self-knowledge') and knowledge of what one will do in intentional action.[13] I emphasize a point that I develop much further in the Dewey Lectures. The point is that whereas non-empirical self-knowledge is not subject to brute error, knowledge of what one will intentionally do *is* subject to brute error. I also

---

[13] See Bernard Kobes, 'Mental Content and Hot Self-Knowledge' in M. Hahn and B. Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.; MIT Press, 2003), 201–228.

emphasize that not all non-empirical self-knowledge is knowledge of active or action-motivating states. Some non-empirical self-knowledge is of passive psychological states, like memories or standing perceptual beliefs. Here is a respect in which the performative aspects of Descartes' *cogito* cannot model all instances of non-empirical self-knowledge.

Both Kobes and I (section II) highlight the role of preservational mechanisms in the non-empirical warrants for types of self-knowledge that go beyond basic self-knowledge. I do not, however, take all instances of non-empirical self-knowledge to derive, by way of such preservational mechanisms, from instances of basic self-knowledge. I think that some non-empirical self-knowledge and the psychological states that are self-known are irreducibly non-active. Some is irreducibly knowledge of non-performative mental states. In this view, I follow Spinoza, not Descartes.

One other large theme in this essay is that warrants for non-empirical self-knowledge are immediate—non-inferential. This point is, of course, another one that Descartes highlighted regarding *cogito*.

Even though it is immediate, the self-knowledge that is illustrated in the pure, self-verifying instances of *cogito* is warranted by justification. (A pure instance is illustrated by an actual thinking of a thought that has the form I am hereby thinking [entertaining] the thought that snow is not white.) That is, the proposition, as thought on the occasion, is self-evident in the sense that it constitutes a reason for itself. Indeed, the self-verifying character of pure instances of *cogito* makes such thoughts infallible. If someone thinks a thought of that form, the thought is guaranteed to be true. There are also impure *cogito*-like instances of self-knowledge. For example, a performative judgment of the form I hereby judge that writing requires concentration is impure inasmuch as it is not self-verifying by virtue of its form: instances can be false. I could judge that I am hereby judging that writing requires concentration, but not in fact do so. I could be distracted, so that I did not in fact engage in the judgment that I judge that I am making. Such errors are pathological, but possible. Although the form of the judgment is not strictly self-verifying, I am inclined to count the occurrent thoughts as reasons for themselves in cases where they are warranted and the individual's understanding encompasses not only the content but an appropriate, performative (as opposed to purely descriptive) use of the content. Thus I think that self-knowledge of this sort is warranted through a justification—as pure cases of *cogito* are.

By contrast, non-*cogito*-like, non-empirical instances of self-knowledge are warranted as entitlements. For example, a judgment I believe that I am older than my sister is not in any sense a reason for itself. There is nothing here analogous to the performative make-itself-true aspect of the *cogito*-like cases that suggests that the nature or content of the belief supports its own truth. In my reply to Kobes, I emphasize that most instances of self-knowledge are of this non-make-itself-true sort. They are non-inferential, but they are not reasons for themselves. The warrants for these cases of self-knowledge do not lie in justifications. They reside

in the reliable, minimal, competence-understanding that fits within a rational structure, such as the structure of critical reasoning. The preservational mechanisms mentioned earlier are the bedrock of this structure. I do not elaborate this point, however, until 'Self and Self-Understanding'.

I said at the beginning of this Introduction that I use warrants for non-empirical cognition as a way of investigating cognitive powers distinctive of persons. Human beings are the only persons we know of. My interest in the distinctiveness of persons is, of course, not driven by interest in the biological distinctiveness of human beings. I want to understand powers, governed by basic cognitive and practical norms, that mark human beings as psychologically special, indeed, I think, especially valuable. The cognitive norms are associated with critical reason. Critical reason is the power to engage in meta-reasoning that marshals reasons and evidence, and that checks and reviews itself in the face of actual or possible criticism. It is the sort of reason that underlies the empirical and mathematical sciences. It drives philosophy itself. Practical norms for critical reasoning cluster around the notions of freedom and responsibility, and are epitomized by moral norms. Critical reason plays an essential role in the powers that underlie these practical norms.

Individuals that have the psychological powers to be subject to these norms are traditionally called 'persons'. Human beings may not be the only persons. Perhaps there are other critically rational beings capable of critical reason and morality in the universe. I am interested in the psychological powers that make persons persons.

The Dewey Lectures, given at Columbia University in December 2007 and substantially revised for the 2011 publication, explore a key element among such psychological powers—the power of self-understanding exercised in non-empirical self-knowledge. Given that self-knowledge is constitutively necessary for being subject to norms of critical reason and morality, I ask what psychological powers, and what psychological and normative structures, underlie this constitutive necessity. The answers provide an outline of a theory of the psychological and normative bases for non-empirical self-knowledge.

'Some Origins of Self', the first of the lectures, begins by discussing notions of person and self. I take *persons* to be something like what Strawson took them to be—individuals with both physical properties and certain distinctive psychological capacities. Very young children and demented adults lack the relevant psychological capacities—centrally, capacities for critical reason and for moral responsibility. They are persons by virtue of being the kind of individual that *naturally* has the relevant psychological powers in mature, undamaged states. I take *selves* to be individuals, or phases of individuals, that actually have the psychological capacities that make persons persons. Kant characterized persons in a way that comes close to capturing what I mean by 'self'. He followed Locke in centering on a capacity for self-consciousness with a diachronic reach—a capacity to be conscious of oneself as oneself, and as one is at different times.

Most of this first lecture consists in reflections on phenomena that occur in the prehistory of two features of selves—self-consciousness and a certain sort of memory of one's own psychological past. This discussion should be construed as a detailed development of transitions from simple versions of ego-centric, or *de se*, indexes, which I discuss in 'Memory and Persons' and '*De Se* Preservation and Personal Identity: Reply to Shoemaker' (both in Part III of this volume), toward a full first-person concept.

The three milestones that I discuss in the prehistory of the relevant sort of self-consciousness are passing the mirror test, imitation, and joint attention. In each of these cases, I discuss crossmodal (visual/proprioceptive) and intramodal (visual) couplings of ego-centric indexes—a pre-conceptual representation that is a precursor of the first-person concept I. I try to show how these psychological couplings provide pre-conceptual precursors that are psychological bases for the eventual capacity to coordinate first- and third-person conceptual points of view on oneself. I believe that this latter sort of conceptual capacity is part of the minimal basis for having a mature self-consciousness capable of objectification. Such a capacity is expressible in uses of the mature self-concept that enters into critical reason and moral responsibility.

In the latter part of the lecture, I discuss various types of memory that are known to occur in human psychology. This taxonomic work leads up to what I call '*meta-psychological, autobiographical episodic memory of one's past from the inside*'. The moniker is, of course, a mouthful. But being careful about the exact nature of the memory pays dividends. For it is a type of memory that is central to the most important psychological kinds. I believe that this is the type of memory that Kant had in mind in his characterization of persons (selves). I think that the same type of memory was what Locke had in mind in his famous attempt to explain in terms of memory the identity of persons over time. Locke was trying not only to produce an account of the constitutive diachronic essence of persons. He was also trying to explain a necessary condition on moral responsibility.

Locke fails to provide an acceptable account for either purpose. I point out relatively obvious (and frequently cited) difficulties with his accounts. In particular, I think that in his account of moral responsibility, he takes memory to have a role at too specific a level. One can be morally responsible for acts that one cannot at all remember engaging in.

Darwin highlights memory in his remarks on wherein human beings are distinctive. I point out relatively obvious difficulties with his account as well. I believe, however, that Kant, Locke, and Darwin are all onto the fact that a certain type of memory—I think, *meta-psychological, autobiographical episodic memory of one's past from the inside*—does figure centrally in our conception of selves, and our conceptions of critical reason and moral responsibility. In the subsequent lecture, I try to show what this role is. This first lecture concludes by forecasting that this and other types of memory figure centrally in a kind of self-understanding that is partly constitutive of selves.

In 'Self and Constitutive Norms' I expand the methodological strategy outlined in 'Our Entitlement to Self-Knowledge'. I assume that a minimal type of self-understanding is constitutive to being a self. By reflecting on the applicability of two types of norms—those of critical reason and morality—I try to make progress in understanding the particular sort of self-understanding that is central. I think that certainly the first of these two norms, and arguably the second, are constitutive to being a self. By reflecting on the nature of the applicability and applications of these norms, I hope to gain insight into the nature of the relevant self-understanding, and ultimately into the nature of selves and persons.

In the course of exploring the applicability and application of these norms, I develop a notion of an *apperceptive core point of view*. Apperceptive core points of view are the aspects of self-understanding that set standards for successful realization of the relevant norms. In the moral case, they also set a baseline for mitigating negative evaluation. Apperceptive core points of view are starting points for inferences that are to be evaluated or that figure in evaluations.

After motivating and developing these points, I return to *diachronic* powers of self-consciousness. I derive the conclusion that meta-psychological autobiographical episodic memory from the inside is constitutive to being a self. I believe that this derivation constitutes substantial progress on Kant and Locke's remarks about the constitutive role of diachronic self-consciousness in being a self.

I carry out such a derivation with respect to three diachronic constitutive features of selves—elements in ordinary inference, elements in practical decision-making, and dialectical elements of both critical reasoning and moral thinking. For example, in the case of ordinary inference, I carry out the derivation from two assumptions about selves. First, I assume that selves, critical reasoners, and moral beings must be capable of propositional inference. Second, I assume that selves, critical reasoners, and moral beings must be able to think, meta-psychologically and consciously, about their own propositional inferences, as their own.

In 'Self-Understanding' I give a unified account of how immunity to brute error in self-knowledge is constitutively necessary for the applicability of norms of critical reason and morality. A large subspecies of the relevant type of self-knowledge—the subspecies that does not involve knowledge of non-representational sensory states—is non-empirical.

I emphasize that except in *cogito* cases, the self-attributions whose warrant guarantees truth *cannot* be determined, or shown to be thus warranted, by appeal to the contents and modes of the attributed states. I argue, however, that constitutively for selves, there are some cases of immunity to brute error. These are the cases that ground applicability of moral norms and norms of critical reason.

The self-knowledge that is integral to critical reason and morality is not deep. It contrasts with wisdom, which we admire and is hard to come by. If the self-knowledge integral to critical reason and morality were deep, it could not be a

condition on being critical reasoner. There are many unwise critical reasoners that have not used their critical reason very well. If it were hard to come by, we would admire it more. For all that, the relevant self-knowledge is warrant-factive, or immune to brute error. And it is philosophically important because it is constitutive to being a critical reasoner, and to being a moral agent.

These points are intuitively plausible. I argue for them by appealing to the base-line conditions for evaluations of critical reasonability and morality. If an individual's self-understanding of a psychological element can be warranted but mistaken, the individual's self-understanding is too detached from that element to count as understanding it from the inside. In such cases, the individual is more fundamentally accountable for the meta-representational attitude than for the self-understood psychological element. The relevant norms require that the individual be able to understand what he or she does or undergoes *in* the doing or undergoing. Such understanding lacks the contingent relation between subject matter and point of view that makes for liability to brute error.

The key underlying idea is that immunity to brute error resides in use of the capacities whose structures and natures are constitutive to a type of point of view. The relevant capacities are mostly preservational capacities—such as inference or purely preservative memory or meta-level redeployment of concepts in speci-fying themselves—that hold a point of view together. I think of these preserva-tional capacities as fundamental structural components of a point of view. They are like the skeletal transit lines of a transportational system, except that the system here is a representational psychology. The main types of these skeletal lines make a point of view, and an individual with the point of view, the kind of point of view, and individual, that it is. More complex psychologies, and individ-uals, are marked by special types of preservational capacities. For example, the preservational capacities in self-knowledge, communication, and reflection mark human beings as cognitively special—and different in kind from other types of individual.

Thus, the discussion of preservative capacities in this lecture, 'Self-Understanding', should be understood in the context of the discussion of purely preservative memory that I first introduced in 'Content Preservation' (Part II). Preservation in communication and purely preservative memory in inference are instances of a deeper preservational phenomenon. My discussion of the role of preservative capacities in this lecture is my fullest one. It best elicits the consti-tutive role of such capacities in determining the natures of basic types of individuals with representational psychologies.

In order to understand immunity to brute error in self-knowledge, I consider immunity to brute error in three other cases: (a) warrant for first-order deductive inference; (b) warrant for beliefs in simple, self-evident truths on the basis of understanding them; and (c) warrant for non-inferential beliefs that one attitude (or content together with mode-type) is a reason to believe another. I use these comparison cases to provide a context for reflecting on the warrants involved in

the relevant type of self-understanding. I see the most basic warrants as attaching to certain preservational powers that are basic to any representational mind. These warrants combine with relevant meta-representational competencies to mark selves as distinctive.

Here I discuss again the way in which the role of the relevant type of self-knowledge in critical reasoning bears on the warrant for the self-knowledge. This is the issue raised by Peacocke with regard to 'Our Entitlement to Self-Knowledge'.

One is warranted in non-empirical self-knowledge through understanding relevant self-attributions. To be epistemically warranted, the understanding of these self-attributions must be reliable in yielding veridical cognition. There are three key aspects to this understanding of self-attributions that connect to critical reason. Each of these aspects is shared with critical reason and is psychologically more basic than—hence independent of—the self-knowledge itself.

First, all instances of non-empirical self-knowledge—including both pure *cogito* instances and self-attributions of standing beliefs—depend on inter-level preservational structures. These are structures that transmit warrant between object-level and meta-representational levels of cognition. In all instances of non-empirical self-knowledge, these structures are shared with critical reasoning. Fulfilling epistemic norms associated with use of these preservational structures is shared between the relevant self-knowledge and critical reasoning. These structures are not more fundamental to self-knowledge than to critical reason. For they are equally basic to inter-level reasoning and inter-level self-attribution. So warrants for relying on these structures in self-understanding are the same warrants for relying on these structures in critical reasoning. The understanding involved in self-knowledge and the understanding involved in critical reasoning have a common source—which is not proprietary to self-knowledge—in the minimal understanding involved in making transitions along these preservational routes.

To summarize the point: There is a set of inter-level structures, which include as subspecies both purely preservative memory and use of object-level contents in meta-level specifications of them, that are constitutive to a mature first-person point of view. These structures and the understanding necessary to make use of them in thought are constitutive to both non-empirical self-knowledge and critical reasoning. The relevant understanding is not more basic to the warrant for self-knowledge than it is to the warrant for critical reasoning. So one's warrant for non-empirical self-knowledge gains some of its authority from an aspect of understanding that is constitutive to critical reasoning and self-attribution, and not more basic to one than the other.

Second, the aspect of the self-understanding that consists in understanding the first-person concept constitutively involves recognizing and implementing responsibility for supporting one's first-level propositional attitudes with reasons,

where one can.[14] Part of implementing responsibility for rationally supporting one's first-level propositional attitudes is mastering the rationally immediate inter-level application of reasons.[15] This recognition of the role of the first-person concept in marking responsibility for implementing reason support for one's own attitudes is part of what is distinctive to the understanding involved in critical reason. It is also part of the understanding of the first-person concept. Hence it is part of the understanding that warrants self-attributions constitutive of self-knowledge. Here the force of the warrant for self-attribution comes from critical reason, not a more basic common source.

To summarize the point: using the first-person concept in a knowledgeable way constitutively involves the understanding of reasons employed in critical reasoning. Whatever warrant such understanding provides partly depends on warrants associated with the understanding of reasons.

Third, the understanding employed in the self-attribution in non-empirical self-knowledge constitutively includes a capacity to specify the self-attributed contents in a way that relies on thinking and understanding those contents *in* their specification. This specificational capacity depends on a first-level understanding of relevant reasons for or against commitments to the relevant content. Understanding a content involves understanding reasons that support or undermine attitudes with that content. This understanding takes a meta-representational form in critical reason. One not only understands *how to* marshal reasons for or against a content. One also understands such reasons as reasons. Insofar as non-empirical self-knowledge figures in critical reason, the self-understanding that generates warrant for the self-knowledge is necessarily enhanced by this understanding of the way any given self-attributed content is embedded in a network of possible supporting- or undermining reasons. This point—made most fully at the beginning of section VII of 'Self-Understanding', the third Dewey Lecture—elaborates the remark in 'A Century of Deflation and a Moment about Self-Knowledge' that a person is an authoritative judge regarding some of the attitudes he or she has because persons "know the law". They know the rational norms and rational connections that help individuate the attitudes that they are self-attributing. They are more reliable at specifying and self-attributing contents insofar as they can locate the contents in a network of reasons. For a position in such a network is an important aspect of the very identity and individuation of the content. So the warrant for the self-understanding that underlies non-empirical self-knowledge gains some of its force from powers of understanding that are constitutive to and distinctive of critical reason.

To summarize the point: The warrant for non-empirical self-knowledge in critical reasoners gets some of its force from the way a critical reasoner must use

---

[14] I emphasize these points mainly in 'A Century of Deflation and a Moment of Self-knowledge' and in 'Reason and the First-Person'.

[15] I am referring to the inter-level structure discussed toward the end of 'Our Entitlement to Self-Knowledge'.

understanding of reason relations, as such, in identifying and canonically speci-fying the representational contents of his or her attitudes.

In the latter sections of 'Self-Understanding', I sketch a relatively full account of non-empirical self-knowledge—the sort of account merely intimated in 'Individualism and Self-Knowledge'. Specifically, I extend the account of authoritative self-knowledge beyond knowledge of one's occurrent propositional attitudes to knowledge of one's standing attitudes (such as beliefs), to meta-representational episodic memory, to anticipations of actions, and to certain non-propositional psychological states, including one's sensations.[16] In all these cases, I take the non-empirical warrant for the relevant self-knowledge to consist in the understanding involved in thinking the known contents.

I hope that the fuller account vindicates my emphasis in 'Individualism and Self-Knowledge' on instances of *cogito* as keys to understanding authoritative self-knowledge more generally. Although *cogito* cases are very special cases, they contain, in germ, many of the key features that are central to and constitutive to all types of authoritative self-knowledge. *Cogito* instances are special and peculiar in their self-evidence, self-justification, and self-verification. But they point beyond themselves in other respects: their non-inferential immediacy; their first-person character; their use of canonical specification of contents that requires understanding of the referred-to contents as well as the specification; their use of a betokening understanding of the attitude mode; their reliance on inter-level representational relations that are routes for preserving warrant; their immunity to brute error; their being warranted through understanding.

The essay ends with a contrast between *de re* understanding—immediate, non-inferential understanding of particulars—and generalized understanding of laws or essences. Self-understanding is a prominent case of *de re* understanding. I think that being a self, and being a person, rest as much on understanding *particulars* in ways that are immune to brute error as they do on uses of reason and generalized understanding that are immune to brute error. Although the latter loom larger in the history of philosophy, the former—as Descartes realized—are equally important for understanding cognition that is distinctive of persons and selves.

II

Part II centers on cognition that derives from interlocution—linguistic communi-cation with others. Communication lies at the heart of most of our knowledge that goes beyond perceptual belief. It is central to what makes possible a shared culture and a progressive science in a complex world. Oddly, it has not been

---

[16] Knowledge of one's sensations is empirical. However, this form of self-knowledge shares important features with the non-empirical cases—especially the immediacy, the direct role of the referred-to state in understanding the self-attribution, and the immunity to brute error.

very central in the history of epistemology. It has been treated as a specialized topic—more a curiosity than the hugely central matter that it actually is.

My discussion of interlocution is part of systematic reflection on the higher powers of mind—powers of intellection. Relations between self-knowledge and reflection, on one hand, and interlocution, on the other, connect closely to relations between first-person and third-person specifications of psychological matters.

The reader should not allow the compartmentalization of the essays in this volume into the four parts to obscure these relations. I have noted earlier the important connections between 'Content Preservation' and some of the essays on self-knowledge in Part I. 'Content Preservation' is also closely related to 'Reason and the First Person' in Part III. To understand much of what goes on in the volume, one might do well to read 'Content Preservation' first.

'Content Preservation' introduced two notions that I think important for epistemology and philosophy of mind. One is purely preservative memory. To repeat my earlier explication: *purely preservative memory* is a type of memory that functions to preserve past attitudes, with their cotents, for current use, without introducing new contents or attitudes (for example, as premises), with their own warrants and subject matter, into current cognition.[17] This notion figures not only as a model for understanding communication in 'Content Preservation'. It also figures in the account of self-knowledge in the Dewey Lecture 'Self-Understanding' and the account of the unity of persons in 'Memory and Persons' and '*De Se* Preservation and Personal Identity'.

The other notion that I introduce, and that I will focus on first, is entitlement. To repeat my earlier explication: An *entitlement* is a warrant whose force does not consist, even partly, in a reason. An *epistemic* entitlement derives from an individual's meeting an epistemic standard for using a good route to truth, where the standard is not that of having a reason. A *practical* entitlement derives from the individual's meeting a practical standard for operating well in the service of realizing some purpose or practical function, where the standard is not that of having a reason.

Entitlement is one side of the distinction, cited earlier, between two types of warrant. The other side, justification, is traditional. The notion of entitlement, isolated as applying to a distinct type of warrant, is relatively new. Attempts to understand warrants purely in terms of individuals' reasons inevitably hyper-intellectualize warrant.

Hyper-intellectualization emerges vividly in accounts of the some of the most basic types of knowledge. For example, the idea that young children (say, under the age of three) or higher non-human animals either do not know anything or must have reasons for their perceptual beliefs to have knowledge through them is not credible. A reason for a perceptual belief would have to be a propositional content that makes reference to perception as such. There is evidence that very

---

[17] In 'Content Preservation', I explicate the notion purely preservative memory in terms of preservation of thoughts and their contents. Here I broaden the explication to include non-propositional states and their contents.

young children and higher animals do not make reference to their perceptual states as such. They certainly do not refer to them, even unconsciously, in propositional attitudes that operatively sustain or support their perceptual beliefs. Yet excluding young children and higher animals from the circle of knowers seems narrow and provincial.

Even if it turns out that children can refer to perceptions as such at very young ages, it seems clear that being warranted in having perceptual beliefs does not depend on having a reasoned justification that cites the belief's deriving from perception as a ground for the belief. It is enough that the belief does in fact derive from perception in the right way. An individual need not know that it does.

Some of the over-concentration on justification, and the neglect of entitlement, derives from a tendency in the history of philosophy to conflate ordinary knowledge and warranted belief with *scientia*—a type of knowledge grounded in a deliberative, self-conscious methodology, as epitomized by science. Much of the history of philosophy is, understandably, centered on *scientia*, not on ordinary, relatively unsophisticated knowledge. Still, the systematic neglect of the less sophisticated types of knowledge distorts epistemology. Indeed, it ultimately distorts the epistemic ground for empirical *scientia*.

Similarly, our reliance on inference—in induction or deduction—and on general cognitive capacities such as perception and empirical belief formation is fundamentally warranted not through meta-justifications, but through entitlements. Eliminating philosophy's long-term tendency to hyper-intellectualize warrant seems to me to be a long-term project that requires the cooperation of many minds to be successful.[18]

Some subtle issues in some of my previous explications of the notion of entitlement need to be noted here. I write, '... entitlements are epistemic rights or warrants that need not be understood by or even accessible to the subject.... Justifications ... involve reasons that people have and have access to'.[19] Although the formulation is not explicit, other writing in the context makes it clear that to have a justification an individual need not have access to a reason considered *as* a reason. Animals and very young children do not have the concept reason and cannot think of anything as a reason. They *can* have reasons—justifications—inasmuch as they engage in propositional inference. I think that higher animals and young children do have reasons for some of their beliefs and activities. What I doubt is that they have reasons for their perceptual beliefs, and more generally that being warranted amounts to having a reason. The reasons that they have use perceptual beliefs as premises, as support for other beliefs. They

---

[18] I do not mean to suggest that the introduction of the notion of entitlement is the only current in philosophy that opposes the endemic hyper-intellectualization in epistemology. Quine's attempt to naturalize epistemology and various reliabilist programs point in the same direction. I think that many of these other programs underplay the role of norms that psychological states and processes must meet in order to contribute to knowledge. I think that the notion of entitlement is important for understanding the normative standards that underlie knowledge.

[19] 'Content Preservation', *The Philosophical Review* 102 (1993), 230.

lack reasons for their starting points—the perceptual beliefs themselves. They are nonetheless epistemically entitled to those beliefs.

The subtlety that I believe that I was insufficiently sensitive to when I introduced the notion of entitlement concerns the notions of having a reason and having access to a warrant. The initial explication of the notion of entitlement in 'Content Preservation', just quoted, centers on access. If access is understood in one common way, modular states are not accessible to an individual, even though the states are in the individual's psychology. The individual cannot bring them up, even on reflection, even under psychoanalysis, just by thinking hard. On such an understanding, if there is modular reasoning, the reasons involved would count as entitlements rather than justifications. I think that it is legitimate to understand *an* entitlement–justification distinction in terms of access—accessibility to being made conscious through hard thinking, without learning new information.

I came to think that a functional-structural conception is more basic than a conception in terms of accessibility to consciousness. As I now use the terms, *justification* consists in *warrants by reasons*; an individual has a justification only if a relevant reason is *present in an individual's psychology*; *entitlements* are *warrants that do not consist, even partly, in reasons*; an individual can have an entitlement even if no rationalizing explanation of the entitlement is present in the individual's psychology. As always, I take reasons to be propositional. If there are modular reasons, they are justifications in an individual's psychology, although not justifications accessible to the individual's consciousness. The justification is accessible only in the weak sense that it occurs in the individual's psychology.

The terminological issue is not fundamentally important. What is basic is being clear about the different ideas involved: the form and function of the warrant in the individual's psychology, and the accessibility to consciousness of the warrant.

The other notion introduced in 'Content Preservation' is purely preservative memory. When memory, long- or short-term, functions as purely preservative, it functions to retain the mode and representational content of a past psychological state, without introducing that content (with its mode) as contributing a new premise with its own warrant into an inference, or as constituting a new topic in the cognitive context.

Recall that lines of preservation form the skeleton of a representational psychology. Types of preservation are a large factor in determining the *nature* of any reprsentational psychology. 'Content Preservation' is the first place where I introduce the most basic type of preservation—purely preservative memory. Although I emphasize the role of purely preservative memory in propositional inference, it takes different forms. In one form or another, it is one of the functions of memory in any representational psychology. Thus, for example, preservation of perceptual contents in memory is a type of purely preservative memory.

The notion of purely preservative memory is important for understanding inference. In any inference, past steps must be redeployed in the inference. When they are reinvoked, they must have been preserved by memory from the

time when they were first invoked. If the reinvocation required a new warrant, one could never rely on the previously invoked step. One would need to introduce a new step with its own warrant. Such a procedure would yield an infinite regress. To understand the epistemology of inference, one must conceive the role of memory as that of an enabling condition, not a source of warrant in the inference. One can be warranted (by entitlement) in relying on memory, just as one can be warranted in relying on a capacity to infer according to an inference rule. The memory does not, however, figure in warranting any step in the inference, even steps that involve reinvoking past steps and thus that rely on memory.

This point is relatively obvious when applied to inference. I think that the interest of the point lies in isolating this particular preservative function for memory—a function that is central to understanding the psychology and episte-mology of a wide range of cases that go beyond obvious cases of inference. Preservation over time of the mode and content of psychological states is a (I am tempted exaggerate and say 'the') fundamental unifier in a representational psychology. As noted in my discussion of essays in Part I, my account of self-knowledge makes central reference to this type of preservation.

The primary objective of 'Content Preservation' is to develop a conception of interlocution (commonly called 'testimony') that emphasizes the unreasoned primitivity of passage of information through communication. I explore the idea that a very primitive entitlement for accepting the word of others is non-empirical. The entitlement is often overridden or overdetermined by consider-ations that center on the particular context, content, or author of a piece of communication. Still, default acceptance of the word of another, other things equal, is an epistemological starting point. I believe that this point fits our actual practice, when that practice is examined in a careful way.

Thus although the idea that we have a non-empirical entitlement to accept what we comprehend as being told can seem to be an unusual idea, it seems to me to apply quite naturally to our actual practice—if one is careful to distinguish general aspects of the practice from aspects that are particular to occurrent communications. These latter aspects can affect and ultimately determine the nature of our warrant (or lack of warrant) for accepting what others say. Such aspects may qualify the default entitlement—either by raising warranted doubts, or by supplying further empirical support. The warrants that derive from particu-lar features of a given communication are usually empirical. I believe, however, that we are entitled to the actual bias that we have—as young children and as adults in unloaded communicative situations—toward acceptance-as-true of what we understand others as presenting-as-true, other things equal. I believe that the default entitlement to this bias is apriori—independent from sense experience for its warranting force.

The main point of the essay is to explain the entitlement to the bias. The idea, stated very summarily, is as follows. The intelligibility of a propositional asser-tion is apriori related to the assertion's having an origin in a being with reason. For propositional capacities are constitutively associated with propositional uses,

which include uses as reasons. Such reasons either support or sustain belief. A function of reason is to support and sustain truth impersonally. So reasons are reliable supporters and sustainers of truth—including sincere, communicative, reliable supporters and sustainers—other things equal. To have a capacity with a function, an individual must be competent, other things equal, to realize the capacity. Any failure to realize the capacity can be assumed to be an aberration, which could not have figured in the normal conditions for obtaining the competence. So a reasonable being is, other things equal, a reliable source for truth. Unless one has reason to think that the conditions of a particular communication act are not normal conditions, one is entitled to rely upon their being normal. So one can rely on what a being with reason says unless one knows something that countermands such reliance.

I further argued that although understanding a communication event as an intelligible propositional assertion always requires the use of perception, in some cases (cases of understanding that do not involve perception-backed *de re* cognition—cases of what I call 'intellectual understanding') perception figures as an enabling condition rather than as a contributor to the warrant for the understanding. For example, to be intellectual understanding, the understanding does not involve seeing some particular that the interlocutor is pointing to. So I limited my claim to utterances of truths such as <u>cats are smaller than elephants</u> and <u>2 is a prime number</u> that do not involve perceptually-backed *de re* applications.

Here I used the analogy to purely preservative memory as it functions in inference. I maintained that perception can function as a trigger for understanding, as perception of diagrams in mathematics often does. I held that perception can function to bring into operation an understanding that is warranted intellectually.[20] There are differences between the mathematical case and the case of utilizing understanding of what is said. I discuss these in both 'Content Preservation' and 'Interlocution, Perception, and Memory'.

The idea of an understanding that is causally dependent on sensory-causal relation to what another says, but that is not warranted through sense perception or through the nature of the sensory registrations, is obviously a delicate one. The understanding must be reliable to be warranted. So the causal chain relating the individual's understanding to the psychological states of the interlocutor must be reliable, and the chain must go through sensory channels. I wanted to emphasize

---

[20] In 'Comprehension and Interpretation' in L. Hahn (ed.), *The Philosophy of Donald Davidson* (Chicago: Open Court Publishers, 1999), I criticize the view that understanding in linguistic interchanges is fundamentally a matter of interpretation—which consists in an inference to the best explanation from perception of linguistic events that provide the basis for the inference. In such an inference the perceived linguistic events are initially construed as non-contentful, or at least as lacking specific content: the content must be inferred by the individual from those events together with background assumptions about the most plausible meaning. It may be that some analog of such inferences occurs in comprehension. The process that leads from linguistic perception to comprehension is certainly a complex one, with many transitions. I believe, however, that it is important not to assimilate this process to person-level inference. It is sub-personal, and it may well be, in major respects, non-propositional.

that apart from the sheer reliability of the sensory channels in connecting to other individuals' thoughts, nothing about the nature of the sensory registrations or the perceptual representation, is essential for the understanding's being warranted. I claimed that although the understanding must make use of a causal-sensory channel, the causal relation is an enabling condition, not a contributor to the warrant.

I now believe that these claims in the "further argument" were mistaken. I believe that human beings do not have a non-empirical warrant for their comprehension of what others say. I stand by the claim that we have an apriori entitlement to rely on the Acceptance Principle. I believe that we do have an apriori default (overridable) entitlement to believe what we understand others to be telling us. Contrary to what I maintained in 'Content Preservation' and elsewhere, I think that the comprehension that is needed to bring pieces of communication from others under the Acceptance Principle is inevitably warranted empirically. The force of one's warrant for one's comprehension depends on perceiving others' linguistic output competently and reliably. In 'Postscript: Content Preservation', I diagnose and criticize my mistake. I try to explain what led me to it, why some replies to it seemed unconvincing, and why it is mistaken.

The mistake was a significant one. It occurs in several of the essays in this section. It affects the argument for an apriori route to knowledge of other minds, sketched in 'Reason and the First Person' (Part III of this volume) and intimated in 'Computer Proof, Apriori Knowledge, and Other Minds'. However, it seems to me that much of what I wrote on interlocution retains value and is unaffected. Despite its defects, the defence of the mistaken claim does, I think, bring out the relatively small role that perception plays in the whole process. The role of perception in comprehension is crucial. In humans, it is ineliminable. But language perception is deeply informed by intellection in its contribution to comprehension. And I continue to think that our default warrant for accepting as true what we comprehend as being told to us is thoroughly non-empirical. Our overall default warrant for belief based on interlocution is empirical, but *barely empirical*.

The epistemic status of belief and knowledge that are derived from interlocution is less important than two other matters emphasized in the essay. The first is the primitivity of the entitlement to accept what others say. The second is the importance in communication of relying on others as rational sources for belief. I believe that 'Content Preservation' opened a valuable way of thinking about a rich set of issues.

'Postscript: Content Preservation' serves as postscript not only to 'Content Preservation' but to the other essays on interlocution that are included here. It centers on three issues. One is the role of knowledge in an antecedent chain of communication in making knowledge by a recipient possible. A second is the Acceptance Principle and its role in articulating a default prima facie entitlement to believe an interlocutor. A third is the epistemic status of the initial comprehension of another's utterance. On the second issue, I defend the position of 'Content

Preservation' against some recent criticisms. On the first and third issues, I criticize my earlier views. I regard the first issue as a relatively minor one, at least in the context of the discussion in 'Content Preservation'. The third issue is much more central. It concerns the mistake that I mentioned two and three paragraphs back.

First, I discuss a counterexample by Peter Graham to a claim in 'Content Preservation' that for a recipient to gain knowledge from interlocution, someone in the antecedent chain must have the knowledge. With some qualifications, I accept the counterexample. I point out ways in which it is not a counterexample to a broader (non-equivalent) principle governing the relation between the recipient's knowledge and knowledge in the antecedent chain that I also state in 'Content Preservation'.

The main value of this discussion, apart from acknowledging and correcting a mistake of mine, lies in my development of the nature of the recipient's entitlement in the counterexample case. This account differs from Graham's. It emphasizes the role of sociological natural kinds in grounding the recipient's entitlement. The role of reliance on sociological institutions supplements and overlays the role of reliance on the interlocutor's rationality in grounding the recipient's entitlements. The former type of entitlement grounds knowledge in the case under discussion, whereas the latter does not. I think that the case illustrates the variety of entitlements that an individual can have on a particular occasion. It also illustrates ways in which epistemic norms are grounded in psychological and sociological kinds.

The second topic of the Postscript is the keystone position of the Acceptance Principle in the structure of warrant for relying on interlocution. I discuss some criticism of taking the principle as basic to the epistemology of interlocution. The criticism underestimates the abstractness of the structure of the relevant default entitlement, and indeed mistakes the very nature of default entitlement. Although the Acceptance Principle is rarely the last word in an adult's warrant for relying on the truth of a piece of communication, it is the first word. And its voice both dominates the structure of additional pros and cons and commonly remains in the mix that constitutes a final warrant. It retains a presence even when other types of entitlements, grounded less in the rational nature of one's interlocutor than in the empirically determinable social role of the interlocutor, overlay it.

The third topic of the Postscript is the epistemic status of the initial intake of a report—the exercise of the recipient's capacity for comprehension. As indicated above, I criticize my view—presented not only in 'Content Preservation' but also in several other essays—that one's epistemic entitlement to rely on one's comprehension of what another says as a correct comprehension can be non-empirical. I discuss the four main ideas that had seemed to me to support this claim. I explain in considerable detail what seems to me to be good about these four ideas, but also why they fail collectively to support the counter-intuitive claim. Fortunately, the error can be isolated in such a way that it affects little else that I have claimed about the epistemology of interlocution. In particular,

the structure of the default entitlements to rely on interlocutors' word, and the function and nature of the Acceptance Principle, remain in place.

'Interlocution, Perception, and Memory' answers objections by David Christensen and Hilary Kornblith to views presented in 'Content Preservation'. Although I still do not accept the grounds that they give for their objections, I accept the conclusion of the objection to my view that belief based on interlocution—in particular the comprehension on which such belief is based—can be apriori warranted.

In retrospect, I value the part of the essay that centers on a distinction between ordinary instruments in science and human beings as interlocutors. I believe that the fact that information is passed through a source (a human interlocutor) that can be expected to use reason makes that source an epistemically different "instrument" than an artifact that has not been programmed to go through rational propositional procedures. I allow that in principle artifacts might have the rational standing of human interlocutors. Indeed, computers that carry out proofs in mathematics have something like that standing. However, ordinary scientific instruments, like thermometers or microscopes or measuring instruments, even those that have linguistic labels on their indicators, are more like amplifications of our perceptual capacities than interlocutors whose propositional output can be expected to be backed by rational procedures. The difference corresponds to a difference between natural meaning—tracking natural events in the world and their correlation with other states and events—and representational meaning, the product of a source with propositional attitudes. There is an apriori connection between having propositional attitudes and truth. The connection between the reading on an ordinary instrument's dial and a natural event in the world is the ordinary causal connection that perception relies upon. However, as the next essay to be discussed emphasizes, some artifacts do function as (artifactual) interlocutors. They are constructed to be resources for reason that provide representational content. That content, and the procedures that systematize it, constitute reasons.

In 'Computer Proof, Apriori Knowledge, and Other Minds', I discuss epistemic issues that arise from the fact that mathematicians rely on computer proofs that are too long and complex for any human to check. For the sake of the argument, I make four assumptions: that pure mathematics can be known apriori; that computers are not autonomous thinkers (thus they cannot be regarded as ordinary interlocutors); that some mathematical propositions can be known apriori, even without knowing a proof; and (from 'Content Preservation') that although commonly empirical, knowledge obtained from interlocution can be apriori. As noted, I now reject this fourth assumption. It can nonetheless play the limitative, for-the-sake-of-argument role that it plays in the essay.

These assumptions serve to isolate the main issue of the essay: Does the fact that a mathematical proof is carried out by a computer, and is too long for a mathematician to check, in itself make it impossible for a competent mathematician to know the theorem of the proof apriori? The primary point of the essay is

to argue for a negative answer. I continue to think that this answer is correct, even though I no longer believe that, strictly speaking, it *is* possible, by relying essentially on a computer, to know a theorem apriori. I think that one's warrant inevitably depends for at least a small part of its force on warrant for one's perceptual beliefs regarding what the computer's outputs are. I believe, however, that the details of the discussion retain epistemic interest.

I consider several ideas that may seem to show that reliance on a computer to carry out a humanly uncheckable proof *in itself* forces knowledge of the proved theorem to be empirical. I discuss the following ideas: (a) that reliance on the computer's reliability as a physical machine can be warranted only empirically; (b) that the source's being a computer rather than a thinker renders empirical any warrant for accepting its outputs; (c) that the difficulty and unsurveyability of the proof, and the need to check whether the computer has the power to carry out such a proof, render any warrant for accepting the theorem empirical; (d) that the need to reidentify the computer as a single source of mathematical outputs requires empirically warranted check.

I argue that none of these ideas is correct. The goal of the argument is to elicit non-empirically warranted mathematical powers that can certify the computer as a source sufficiently worthy of trust to take it to yield a successful proof. I discuss analogies between a great mathematician's non-empirical capacity to appreciate his or her own mathematical power and such a mathematician's capacity to appreciate mathematical power from another source, here a computer. The value of the essay lies, I think, in the details of the discussion—focusing on neglected ways in which our non-empirical powers of reasoning can be applied.

'Comprehension and Interpretation' criticizes uses of Quine and Davidson's theories of translation and interpretation as accounts of our normal understanding of others' utterances. Although Quine and Davidson present their pictures as idealized rational reconstructions—not as accounts of actual linguistic under-standing—they focus on sophisticated inferences from evidence, especially in difficult cases. Quine centers on radical translation of a foreigner. Davidson centers on on-the-fly interpretation of non-standard utterances, such as malaprop-isms and irony. I think that the basic cases for theorizing about comprehension are simpler. They are cases in which we reliably and correctly take in what another says without thinking.

The essay develops the point in 'Content Preservation' that our basic non-meta-representational understanding—our comprehension—of what others say or write is not a matter of person-level propositional inference from evidence. (See note 20.) And it is not warranted through reason or reasoning. It is normally warranted through an entitlement that resides in the reliability of a competence to comprehend another's utterance immediately—without going through propos-itional inference based on beliefs about evidence. I confine this point to standing linguistic understanding—the sort of understanding that does not depend on understanding context-dependent devices, such as devices of *de re* demonstrative reference.

I compare, in some detail, the immediacy of such comprehension to the immediacy of perceptual belief. I argue that both are normally warranted through entitlement, not justification. An individual can comprehend another's utterance without even having the concepts needed to justify the comprehension. For example, a child can comprehend its parents' utterances without conceiving them as utterances and without having concepts of meaning or truth.

The child's entitlement to rely on its comprehension rests on the reliability of the child's non-inferential comprehension competence. Similar points apply to default warrants governing standing linguistic comprehension capacities in adults. The warrants are analogous to the basic entitlement that we have to rely on perceptual belief: this latter entitlement resides in the reliability of our perceptual systems in producing veridical perceptual states and in the reliability of our psychological transitions from veridical perceptual states to veridical belief states.

The essay expresses differences with Davidson on the role of the social in psychology. I hold, and still hold, that it is metaphysically possible for an individual to have beliefs without language. I think that very young children and non-human animals are probably actual examples.

I think that it is metaphysically, though not psychologically, possible for an individual to have language without social relations. And I think that one (metaphysically) could have the concept of belief without social relations. On the other hand, I think that dependence on others is metaphysically necessary for having certain types of beliefs, if one lacks certain types of background information. For example, one cannot believe that arthritis can occur in one's thigh, unless one depends partly on others for one's concept arthritis.

The essay also contains two considerations that were meant to further support the view that our comprehension of others can be, and under very restricted conditions is, apriori. One consideration emphasizes the way that our apriori comprehension of our own thoughts is intertwined with comprehension of others' utterances. The other consideration appeals to the possibility of injecting, or mainlining, one individual's thought contents into another individual's psychology, using brain-to-brain causal processes that do not go through perception or the senses.

As indicated, I no longer accept the conclusion that these considerations were supposed to support. The first consideration states something correct about actual enabling conditions for some first-person understanding of some of one's own thoughts. It does not show that comprehending others' utterances can ever be apriori. I discuss the second consideration in some detail in 'Postscript: Content Preservation'.

The value of the essay seems to me to lie in its evocation of the immediate, non-inferential nature of much linguistic comprehension,[21] and of the similarities

---

[21] I re-emphasize that being non-inferential does not mean not involving processes. *Inference*, as I use the term, is processing that connects propositional contents. Similarly, immediacy does not

between comprehension of others' utterances and perceptual belief about ordinary non-linguistic entities. In particular, the primary warrant for both is an entitlement. I believe that the similarities are closer than I believed at the time I wrote the essay: both entitlements are inevitably empirical. Still, the essay elicits how much of our warrant for relying on computers in mathematics is non-empirical. It seems to me that the mistake is instructive, but the main claims of the essay stand.

'A Warrant for Belief in Other Minds' was written in the late 1990s. The last early version of it is dated 1999. Originally, it was titled 'A Non-Empirical Warrant for Belief in Other Minds'. When I first wrote it, I intended to develop the argument that can be found, in truncated form, in 'Computer Proof, Apriori Knowledge, and Other Minds' (section V) and in the last section of 'Reason and the First Person'. That argument supported and employed positions that I no longer accept. As noted above and in 'Postscript: Content Preservation', I no longer believe that our warrant for relying on comprehension of others' utterances is ever non-empirical. In 2010–2011, I revised the essay to its present form, and revised its title. An alternative revised title would be 'A Barely Empirical Warrant for Belief in Other Minds'.

I believe that the argument brings out a distinctive way of knowing other minds—through comprehending speech—that has been under-appreciated in the history of philosophy. I discuss traditional arguments from analogy and arguments from inference-to-the-best-explanation. I argue that although they are part of the account of how we know other minds, most of them focus on too high a cognitive level to account for the primitiveness of one of the warrants we have for believing in other minds. I think that the depth of our reliance on propositional form as a sign of mind needs far more attention. This essay takes a step in that direction.

III

The essays in Part III center on relations between reason, or reasoning, and being a reasoning individual. The individual may be a self—a critical reasoner, capable of self-evaluation—or just a non-reflective, first-level reasoner. Being an individual requires having some unifying condition. The unifying conditions that I investigate are various types of propositional forms and propositional interconnections. The work in Part III is closely connected with the work in Parts I and II on preservational structures in psychologies of all levels—pre-rational, rational, and critically rational. (The connections are closest with 'Content Preservation', 'Our Entitlement to Self-Knowledge', and the last of the Dewey

---

entail lack of processing. A representational state is *immediate* if it is not the product of an inference.

Lectures, 'Self-Understanding'.) These structures are sources of unity that are constitutive to being egos, thinkers, and selves. The preservational structures discussed in Part III are psychologically more complex than the structures that underlie the ego-centric indexes that designate whole psychological individuals. The types of individuals discussed in this part are at least reasoners, and in some cases selves or persons.

'Reason and the First Person' argues that there is a constitutive connection between use of the mature first-person concept and understanding what a reason is. I begin by considering Lichtenberg's doubt that Descartes' *cogito* needs to use, or can support justification of a use for, the first-person concept. I argue that a use for that concept is guaranteed and legitimated by understanding what a reason is.

I distinguish two aspects of understanding what a reason is. One is understanding the role of the concept reason in evaluating or appraising. The other is understanding the role of the concept reason in moving one to apply the evaluations or appraisals in reasoning—to affect attitudes in accord with reasons. Understanding the concept reason requires understanding this implementational role, as well as the evaluative role. Understanding the implementational role requires an ability to mark conceptually, in particular instances, the attitudes or acts for which rational evaluation of an attitude or act enjoins shaping it in accord with the rational evaluation. The first-person concept marks those attitudes that are subject to rationally immediate implementation.

I explore a sense in which theoretical reason has a practical aspect through its connection to rational agency. A critical reasoner would be deficient if the reasoner lacked the apriori understanding marked by the first-person concept. Use of that concept is underwritten by reason.

As noted, the work in this essay connects to work in 'Our Entitlement to Self-Knowledge' and the second and third Dewey Lectures. It spells out connections between one's right to use the first-person concept in self-knowledge and the use of critical reason. And it spells out the constitutive role of responsible commitment to norms of critical reason and morality that comes with being a self or person. The work in the essay should also be regarded in the light of the extensive discussion in the first Dewey Lecture and in 'Memory and Persons' of ego-representations that are ontogenetic predecessors of the first-person concept. (There is some discussion of these predecessors of the mature first-person concept in the nineteenth and twentieth paragraphs of 'Reason and the First Person'.) The first-person concept is part of the capacity to understand and acknowledge the normative standards—practical and theoretical—that are already present in pre-critical reasoning.

Although written later, 'Memory and Persons' and its postscript, '*De Se* Preservation and Personal Identity: Reply to Shoemaker' are natural predecessors for 'Reason and the First Person'. In 'Memory and Persons', I discuss a traditional issue regarding personal identity. Most of the discussion centers on the fundamental role of ego-centric or *de se* representation in some psychological competencies that are fundamental for any reasoner, including non-critical

reasoners that are not persons. Thus the essay discusses psychological predecessors—pre-personal ego-centric representations—for the mature first-person concept. These predecessors are embedded in unifying competencies and structures that mark specific psychological kinds of individuals at various levels of representational competence.

The essay begins by distinguishing experiential memory, substantive content memory, and purely preservative memory. Experiential memory always has *de se* (or ego-centric) content, marking various grades of psychological complexity. The discussion here is an antecedent of the more complex discussion of couplings of ego-centric (*de se*) markers and of autobiographical episodic memory from the inside, which figures in the first two Dewey Lectures. The discussion of purely preservative memory has its antecedent in 'Content Preservation'.

I argue for a fundamental constitutive role for preservative memory with either *de se* content or *de se* presuppositions in three psychological activities: use of perception in action or to serve other needs; carrying out intentions; and engaging in inference. I show how each of these activities is constitutively inter-related with *de se* markers and preservation of such markers in memory. The first two involve a sensitivity to and privileging of the individual's own needs, aims, and point of view. This sensitivity and this privileging are marked by *de se* representational content of states with those competencies. The third activity, inference, presupposes such *de se* sensitivity and privileging in a more complex way.

I maintain that these three activities are constitutively necessary for being a person, indeed, for being any individual with propositional attitudes. The three activities constitutively depend on memory, with either *de se* content or *de se* presuppositions, or both. No individual with propositional attitudes could lack the sensitivity and privileging involved in these *de se* capacities. Having the capacity is part of what it is to be an individual with propositional attitudes (and more specifically, to be a person). Having these *de se* competencies requires applying them veridically. Veridical applications in memory require sameness of individual over time. Memory, with its *de se* presuppositions of transtemporal agent identity, is a condition on the possibility of an individual's having a representational mind. Memory, with its *de se* presuppositions of transtemporal identity is a condition on the possibility of an individual with propositional capacities. Hence it is a condition of the possibility of being a person.

The foregoing sets the stage for a discussion of the traditional dispute over the Lockean idea that the concept of personal identity can be analyzed in terms of the concept of memory. The dispute is over Locke's apparent approach to explaining personal identity over time in terms of memory. Butler objected that such an explanation is circular, for the concept of memory presupposes that the rememberer and the individual with the past psychological states that memory retains are the same individual. Shoemaker and Parfit proposed a way of circumventing Butler's objection. They appealed to a notion *quasi-memory* that allows connection between an individual and the past psychological states of another individual. They tried to explain personal identity in terms of sufficient psychological

connectedness illustrated by quasi-memory, together with a condition that the connections do not branch in certain ways. This latter condition is introduced to handle imaginable cases of fission in which two individuals are equally psychologically connected to the states of the same past individual.

In 'Memory and Persons' I claim that the approaches of Shoemaker and Parfit merely avoid definitional circularity, and that a deeper sort of circularity—explanatory circularity—remains. The point is that the very sorts of psychological competencies whose connections over time are constitutive to individuals, including persons, are themselves essentially marked by *de se* preservational capacities. One cannot get underneath these capacities by appealing to neutral capacities like quasi-memory. For to be the sort of individual that could have quasi-memory, one must have the *de se* capacities as core competencies. Thus having quasi-memory constitutively requires having *de se* capacities. These points are worked out in some detail for each of the three basic psychological activities—use of perception, intentional action, and inference.

The main value of the essay seems to me to reside not so much in refuting a rather special and peculiar reductionist strategy, albeit a variant on an old and famous one. It resides in delineating apriori knowable, constitutively necessary connections among psychological capacities that are constitutive to representational mind, on one hand, and *de se* representation, on the other. *De se* representation marks sensitivity by the individual to his own needs, aims, and point of view. *De se* representation is a phylogenetic and constitutive ancestor of the first-person concept. It marks something central to what it is to be an individual with a psychology—a capacity in such an individual to privilege, be sensitive to, and mark psychological states and their perspectives as individual's own, by relating them to the individual's own needs, aims, and point of view. Psychological states are not specifiable apart from contents that indicate in a *de se* manner the psychological states as the individual's own. All representational psychological states must be connected to representational psychological states that function to privilege the individual as a distinct being.

Shoemaker responded to the argument in 'Memory and Persons'. Unfortunately, the response misconstrues the argument. The response serves to illustrate the power of my argument, rather than serving to reply to it, much less refute it. In '*De Se* Preservation and Personal Identity: Reply to Shoemaker', I show that a scenario that Shoemaker designs to support his approach actually confirms the line of 'Memory and Persons'. His reply fails to come to grips with my claim that having quasi-memories requires and must be explained in terms of *de se* memory. A detailed discussion of his scenario shows that the individuals in it have ordinary *de se* capacities, and that the additional quasi-memory capacities operate in a psychology that has *de se* capacities at its core. These core capacities are constitutively necessary for quasi-memory, whereas the contrary is clearly not so. Herein lies the explanatory circularity in Shoemaker and Parfit's accounts. There is a circle in the explanation of the constitutive natures of the relevant capacities.

I highlight two basic functions for *de se* capacities. These functions must be present in any representational individual psychology—hence in the psychology of any person. One function is to be sensitive to the individual's own needs, goals, and point of view and to privilege them. The second function is to serve those needs, goals, and point of view through representational powers. Part of what it is to have an individual mind is to be sensitive to one's own needs, projects, and perspectives, and to do things for oneself. Powers to carry out the latter function require, as 'Memory and Persons' shows, having *de se* preservational powers. These powers are unifiers of individual minds over time. They mark and presuppose the identity of the individual.

Since quasi-memory is unspecific in its function as to whether it operates intra-individually or connects one individual with another, it cannot realize or explain the unity of persons, or other individuals with representational minds. No other element in the Shoemaker–Parfit position can do so. A special feature of representational psychologies is that the core psychological states make reference to the individual, or some aspect of the individual, in a way that marks the centrality of that individual's core needs and projects for that individual. Psychologies contain *de se* specifications of the individuals who have them, because those specifications mark fundamental functions and powers constitutive of being an individual with a representational psychology. The constitutive nature of such specifications makes it impossible to account for personal identity in terms of psychological states that do not contain them.

This account provides a de-intellectualized and elaborated counterpart of Kant's unity of apperception. Kant insightfully maintained that psychologies contain powers that function to provide a unifying condition for the psychology. These powers, in turn, provide a unifying condition for the individual with that psychology. Kant also took reflexive representation to be an essential aspect of the unification that interested him most (the unification that marks critical reasoners). Thus he recognized that the unifying psychological processes, such as memory and inference, cannot be specified independently of specification, within the unified psychology, of the individual that authors the psychology. His expressions of these insights focused, however, too exclusively on a very high level of psychological sophistication. Because he was interested in *scientia*—the high-level cognition exhibited in science and moral deliberation by persons and selves—he couched his point in terms of self-consciousness. He really meant *self* here—a reasoner capable of deliberation, reflection, critical reasoning.

The basic point, however, does not depend on a high-level capacity for self-consciousness. *De se* states and *de se* preservational powers are present in any representational mind—even animals that have only perception and no propositional attitudes, perhaps even animals that lack consciousness altogether. These powers take on more sophisticated forms in more complex minds. They re-occur in propositional inference—as I explain in some detail in 'Memory and Persons'. They become apperceptive in the self-consciousness and self-knowledge of persons and selves, as I discuss in the Dewey Lectures. But the basic *de se*

unifying functions go all the way down. Tracing their realizations at different levels of representation is a large part of understanding both the essence and the varieties of representational mind.

'Modest Dualism' deals with the classical mind–body problem. This essay's ontological focus distinguishes it from most of the other essays in this volume. The ontological discussion centers, however, on formal structures that mark capacities to reason—capacities that ground the psychological kinds and epistemic norms that are central topics in the collection.

I begin by emphasizing the importance of doing metaphysics in such a way that it hews closely to what we know, especially in the sciences. The essay criticizes a cultural and professional tendency to assume more reason to accept materialist ontologies than we actually have. I review my argument from 'Individualism and the Mental' against a particular version of materialism—the token identity theory.[22] Then I discuss a less committal version of materialism—compositional materialism.

A paradigm that has served the natural sciences well is to take more complex entities that are studied in the special sciences to be composed out of physical entities that are studied in physics. I emphasize that the applicability of this paradigm for physics, chemistry, and biology always had more basis in the actual kinds studied by those sciences than it does for psychology or the social sciences. I maintain that although the paradigm should continue to be explored in trying to understand the relation between psychology and neuroscience, it has no solid, comprehensive evidence to support it.

The main contribution of the essay is a pair of closely related arguments for the conclusion that the natures, the basic kinds, of states studied in cognitive psychology—kinds whose natures involve representational contents with propositional structures—are not kinds whose instances are material composites of physical entities.

One argument centers on causation. According to the natural sciences, reason relations and propositional structures are not structural features of material composites. The causation by material parts of material composites, operating in their physical relations to one another, must suffice to compose causation by the material composites. Given what we currently know, the causal powers and causal structure of material parts do not seem alone to compose the causal powers and causal structure of causal events that involve rational, propositional structures of propositional states and events. Material causation is not itself rationally structured. Rational structures are causally relevant. So rational causes do not appear to be a mere composite of causes by physical parts.

The other argument centers on constitutive structure, and goes as follows. The physical structure of material composites consists in physical bonds (physical relations) among the parts. According to modern natural science, there is no place

---

[22] 'Individualism and the Mental', *Midwest Studies in Philosophy* 4 (1979), 73–121, reprinted in *Foundations of Mind*.

in physical structure of material composites for rational, propositional bonds or relations. The structure of propositional psychological states and events constitutively includes representational content with propositional structure—the sort of content and structure that can constitute reasons. So propositional states and events are not material composites.

Both key ideas in the arguments are deeply ingrained in the relevant sciences. The idea that the physical world, as studied by the natural sciences, does not have the structure of a text—propositional structure—is deeply entrenched since the advent of modern science. The relation between physical properties and relations, on one hand, and the physical particulars that have them, on the other, is not the relation of predication. Propositional negation and quantification are not to be found in the physical world by the physical sciences. The idea that many psychological states and events, at least among the higher animals, are fundamentally propositional attitudes capable of engaging in reason-supporting relations to other attitudes is fundamental in both common sense and cognitive psychology. The depth of these prima facie opposing commitments in the natural and psychological sciences makes it, at least, prima facie implausible to regard propositional attitude states and events as material composites out of physical states and events.

Rejecting materialist paradigms is not rejecting the dependence of the psychological on the physical. I see no ground for thinking that propositional states and events could ever float free of a physical basis. The point of the argument is to bring out that the relations between the physical and the psychological are likely to be more subtle and complex than simple materialist or physicalist positions propose. I believe that philosophical and scientific investigation of these relations should proceed in an exploratory way. I think, however, that it is well for such investigations to bear in mind that in propositional structure and in reason, we are dealing with something that has a quite distinctive form and function, and that is subject to quite distinctive practical and cognitive norms.

'Epistemic Warrant: Humans and Computers' begins by distinguishing again between justification—warrant by reason—and entitlement—warrant without reason. The essay then discusses warrants for transitions in inferences. I argue that many deductive and most inductive inferential transitions are warranted by entitlement, not by justification (reason). Thus, when premises form reasons for conclusions, they often do so through a mixture of justification and entitlement. The premises are reasons for the conclusions. Hence, they are justifying elements. The individual is, however, often warranted in the inferential transitions not by reason, but by entitlement. Since the full warrant for the inferred conclusion involves both premises and inferential transitions, the full warrant is a mix of justification and entitlement. Since justification occurs if a reason plays any role in the warrant, it is correct to call the conclusion justified.

In such cases, the individual's premises count as reasons, justifications, for the individual's conclusion if and only if the combination of the premises and the inferential rules that are actually relied upon yields a rationalizing explanation of the belief-worthiness of the conclusion. (The conclusion can count as justified

whether or not the transitions are cognized in the psychology in a way that makes the warrant for relying on them a justification or an entitlement.) Thus even many deductive and nearly all inductive conclusion-supporting inferences are warranted through a mixture of justification and entitlement.

I then discuss the likelihood that there are cases in which the conditions on premises of a warranted inference counting as justifications of the conclusion is not met. I center on empirically warranted inferences. I assume that the inference tracks likely connections in nature. So it supports the truth of the conclusion. But the inferential transition does not follow a rule that provides any insight at all into the nature or existence of the connections. The transition is the product of brute evolutionary adaptation. So the combination of the premises and whatever rule governs the inferential transition does *not* yield a rationalizing explanation of the belief-worthiness of the conclusion. Still, the inference is a good route to truth. The goodness of the route is stamped into the inferrer, perhaps, by evolution because it does connect environmental properties in a patterned way. Then, I think, the premises and inferential transition entitle one to the conclusion. The premises do not provide a reason or justification for the conclusion. I believe that this account provides insight into the status of blind, but competent, truth-tracking, associative inference.

I turn to reliance on computers in mathematics. I recap some of what I wrote in 'Computer Proof, Apriori Knowledge, and Other Minds' about human warrant for relying on computers to prove mathematical theorems. I argue that in normal cases of computer proofs, the computers function as proxies for reasoners. I also compare reliance on computers for carrying out a proof with the reliance by an intuitive mathematician, like Ramanujan, on unconscious powers to produce theorems, without carrying out a proof. Whereas warranted reliance on computers inevitably involves some inductive reasoning about the computers' reliability and power, Ramanujan's warrant is likely to have been an unconscious entitlement, analogous to entitlements to perceptual beliefs, except that Ramanujan's entitlement probably derived from non-perceptual sources—sources of "intuitive" understanding.

Ramanujan's practice suggests, what is in fact the case, that there are various ways in which successful mathematics goes beyond deductive proof. I center on some epistemic differences between standard deductive computer proofs and probabilistic computer proofs. I argue that the key epistemic difference between the two types of proofs is not that the latter is inevitably less certain or less strongly warranted than the former, but that the latter gives less insight into necessary connections between premises and conclusion. The type of reason that deductive proofs provide for believing their conclusions provides a deeper understanding of the necessity of the truth of the conclusion, and of the necessary relations between premises and conclusion. In all cases in which a mathematician relies essentially on a computer's mathematical operations, deductive or inductive, the mathematician's ultimate warrant for believing the conclusion of the computer's work is inductive and empirical. Still, in understanding the deductive

transition types that the computer goes through, the mathematician gains a deeper understanding of the necessary ways that premises support conclusion—and a deeper understanding of the necessity of the conclusion.

I sketch three types of reliance on computers in empirical work: Bayesian modeling of empirical phenomena, genetic algorithms in modeling empirical phenomena, and genetic algorithms in discovering purported empirical laws. I argue that only the first type of computer transitions model human reason-giving inferences. Genetic algorithms are not proxies for reasoning. The randomized mutations and recombinations are not analogs of reasoning—inferential transitions that provide reasons, justifications, for conclusions.

These differences do not bear on the power of our warrants for believing the computer's results. Inductive warrants can be equally strong in the different cases. My reflections bear on the relations between the different types of transitions in the computer's processing, on one hand, and corresponding steps in human inferential transitions, on the other. It is possible that genetic algorithms provide approximate models for certain types of human inferential transitions. In such cases, previous steps may enhance credibility of later steps—yielding entitlements to rely upon them—without providing explanatory rationalization of the later steps—that is, without being reasons for them.

The third type of reliance on computers in empirical work, use of computers to discover empirical generalizations or laws, may model a type of inference that sometimes occurs in science. Such inferences to good explanations might support their conclusions in the sense that they make the conclusions reliably more credible. But they fail to provide even a partial rationalizing explanation of the enhanced credibility of the conclusions. Such inferences yield entitlements to their conclusions, all the way down, not justifications. It seems to me certain that inductive (or abductive) inferences in science are not all of this sort. On the other hand, I think that genetic algorithms strongly suggest that some inductive reasoning in science may be more like blind Humean association than like the deliberative evidence-explaining inferences that provide the ideal of reasoned scientific explanation.

Understanding non-deductive inferential connections in science is incomplete understanding. It may well be that there is an irreducible reliance on environmentally, evolutionarily stamped-in associative patterns in our inferential transitions, even in science. We would be entitled to rely on such transitions well before they took a form that provides a rationalizing explanation of conclusions that they warrant. The paper uncovers, I think, several levels of incomplete understanding of inductive inferential transitions that suggest a role for blind entitlement, even in science. Although such transitions cannot yield ideal scientific understanding, they seem to be a necessary tool in the development of understanding that produces scientific reasons.

The essay does not explicitly discuss ways in which these types of inference bear on psychological types of individuality. I think it clear, however, that, along with memory, inferential patterns are the prime unifiers of a higher-level

psychology. By considering different types of warrant for inferential transitions, one clarifies different types of psychological competence. Deliberative, critical, reasoned inference in practical reasoning, common sense reasoning, science, and mathematics helps mark the most sophisticated psychological personality. This high-level nature is inevitably dependent on lower-level, more "animal" types of inferential transition that mark less sophisticated levels of psychological individuality. I think that this essay contributes to thinking about relations among these levels.

IV

The essays in Part IV are more exploratory, more "initial", even than most of the other essays in the volume. They barely scratch the surface of a large and fascinating topic—reflection. They touch on reflection as it occurs in philosophy, mathematics, logic, and sometimes in the empirical sciences. They should be read in conjunction with my discussion of the epistemology of thought experiments in philosophy, particularly the thought experiments that led to modern arguments for anti-individualism.[23] The essays reprinted here highlight apriori aspects of reflection. In other work, I highlight empirical aspects.

It is crucially important, in understanding reflection, that one not assume that all warranted reflection is apriori. Very frequently, reflection employs uncontroversial empirical truths—especially general truths that would not, in common parlance, be counted 'theoretical'. I believe that some "armchair" reflection is one natural, and not *always* inappropriate or fruitless, occupation for philosophers. Some armchair reflection is like the reflection in elementary mathematics that relies on apriori understanding. Much of it involves a mix of apriori and empirical elements. Some of it is fundamentally empirical. Much of my work on reflection has sought to develop insights in the rationalist tradition, while emphasizing that reflection is a much more varied phenomenon, both psychologically and epistemically, than traditional rationalist models presented it as being.

It is also crucially important that one not assimilate even apriori aspects of reflection to conceptual analysis. I believe that very little analysis, strictly speaking, occurs in fruitful philosophical or mathematical reflection. I think that the key to making progress on this complex topic is to think about specific cases of successful reflection, remaining open to their differences from traditional paradigms of analysis (the analysis of the concepts bachelor and knowledge), and from one another.

---

[23] See, for example, *Foundations of Mind*, *op. cit.*, 18, 27–28, 162–179, 291–306; 'Some Reflections on Scepticism: Reply to Stroud' in *Reflections and Replies Essays on the Philosophy of Tyler Burge*, M. Hahn and B. Ramberg eds. (Cambridge, Mass.: MIT Press, 2003).

'Reasoning about Reasoning' is a critical discussion of David Lewis's account of iteration in reasoning about reasons in his book *Convention*. The topic comprises both iterative reasoning about one's own reasons and iterative reasoning about another's reasons. I accept this slight refinement on Lewis's gloss on governing *having a reason*: if a person has a reason to believe something and does not already believe it, the person would incur some deficiency in rationality if the person were to disbelieve it; and the person should be able to come to believe it with only a little reflection—or at least with reflection, however much or little, that does not make use of information that the individual does not already have.

The essay makes the point that one cannot iterate reasoning about having a reason infinitely, if each iteration is inductive or has strength less than 100%. Because there is a real difference between having a reason about a subject matter at a given level and having a reason to believe one has that reason (which is a reason at a different representational level), having reason at the lower level does not guarantee that one has reason at the higher level. This point is especially applicable to transitions from first-order reasons (reasons that are not about reasons) to second-order reasons (reasons about having first-order reasons). If one takes seriously the condition on deficiency in rationality, in the gloss on 'having a reason', the point seems to apply in any iteration from one higher-order level to the next: having a reason at one level does not guarantee that (at the next higher level) one has a reason to believe—or is rationally deficient in not believing—that one has that reason.

In fact, it seems certain that at a sufficiently high level of iteration, an individual *i* incurs no deficiency in rationality if *i* has a reason to believe p, but cannot 'with only a little reflection' come to believe that *i* has reason to believe that *i* has reason to believe . . . that *i* has reason to believe p. Combining beliefs whose warranting strength is less than 100% in a piece of reasoning diminishes the warranting strength of the conclusion.

Since all judgments about another person's reasons are inductive and have strength less than 100%, the case of interpersonal reasoning about reasons adds another possible source of friction that drags against infinite hierarchies of having reasons about having reasons.

Looking back, I think that the value of the essay is two-fold. It cautions against over-idealized theories about having reasons—driven more by logic than by a realistic epistemology. And it insists on distinguishing between levels of reasoning. Some reasoners can reason, and thus *have* reasons, but lack a concept of reason. In insisting on being realistic about what it is to have a reason and in emphasizing level differences, the essay constitutes a criticism of hyper-intellectualization in epistemology.[24]

---

[24] See also 'On Knowledge and Convention', *The Philosophical Review* 84 (1975), 249–255; reprinted in *Foundations of Mind*, *op. cit*., for another critical discussion of Lewis that opposes a different, but related, sort of hyper-intellectualization in epistemology.

I include 'Thought Experiments and Semantic Competence: Reply to Benejam' not because it makes a distinctive contribution, but because it contains in very compressed form many of my criticisms of the traditional conception of reflection. It might serve as an introductory orientation to the other essays on reflection.

'Concepts, Conceptions, Reflective Understanding: Reply to Peacocke' is part of a long, on-going constructive interchange with Christopher Peacocke on the nature and scope of apriori knowledge, an interchange that began in 1993 when he graciously helped host my Locke Lectures at Oxford. This essay highlights several areas of agreement, and explicitly shares an appreciation of the contributions of the great traditional rationalists—Descartes, Leibniz, Kant, and Frege. The essay also brings out a significant area of disagreement. Peacocke develops the classical conception of reflection, most fully present in Leibniz and Kant. According to this conception, reflection consists in bringing to clear consciousness an understanding that is already present in the reflector's psychology. In fact, I think that the classical conception was somewhat more liberal than the conception that Peacocke elaborates. It allowed that reflection puts together an explication for the first time from conceptual components that are present, perhaps unconsciously, in the individual's psychology.

I think that even this liberal version of the classical conception fails to do justice to some very important cases of reflection. A thinker may have an incomplete understanding of a concept and may not have the concepts to provide an adequate explication of it. Leibniz's infinitesimal-based understanding of his concepts of limit and differentiation is, I think, an example.

A central difficulty with Peacocke's account is that it overrates the role of unconscious *but fully formed* conceptions in guiding reflection to definition-like explications of concepts (or terms) that are initially not clearly and distinctly understood. I believe that although such guidance can happen, it is uncommon. I think that most reflection is guided by memory of examples—often socially inculcated—senses of similarity among examples, inductive capacities, and inferential patterns that are either too low-level and specific or too generic to provide satisfactory explications or definitions. I think that explications or definitions are commonly discovered—and enter into the psychology for the first time—through the process of reflection. Reflection is rarely merely a dialectical bringing-to-consciousness of an unconscious, "implicit", but fully formed explicational, criterial conception.[25]

The role of dependence on, and perhaps deference to, others for a fuller conception plays a significant role in everyday life. Sometimes one simply has to learn more from others in order to obtain an adequate, reflective understanding of one's own concepts. Sometimes even the experts, or the whole community

---

[25] See 'Postscript: Individualism and the Mental', 'Intellectual Norms and Foundations of Mind', 'Concepts, Definitions, and Meanings'–all reprinted in *Foundations of Mind*, *op. cit.*, and 'Living Wages of *Sinn*' in this volume.

combined, lack sufficient understanding to provide adequate explication of one's concepts. Mere reflection even by experts will not yield sufficient understanding of what guides, justifies, and explicates one's concepts. I make this point with respect to Newton and Leibniz's respective understandings of key concepts in the calculus.[26] I take this point further in the last section of 'Living Wages of *Sinn*'. I make a similar point elsewhere with respect to reflection on empirical concepts. Dalton's understanding of his concept atom and erstwhile communal explications that situated tomato under the super-ordinate concept vegetable are examples.

The essay concludes with a detailed discussion of the relation between non-inferential understanding of simple, self-evident logical truths and the inferential mastery that is necessary for having such understanding. Our justification for believing simple logical truths does not derive from inferring them from premises that are more basic, or better justified, than the simple logical truths. The justification seems non-inferential, and resides in understanding the truths themselves. On the other hand, as Frege emphasizes, understanding the truths requires a background of inferential capacities. One could not understand the truths unless one could connect them to some other truths via inference. One can both infer the simple truths as conclusions and use them to infer to other truths. Such inferential capabilities are essential to understanding the truths in such a way as to be non-inferentially justified in believing them. I explore this combination of ideas through reflection on Frege's account of his axioms.[27]

'Reflection' was written for a conference in Fribourg, Switzerland, in summer 2004. I gave the paper at the conference as a public lecture, but never submitted it for publication. It begins with a review of criticisms of classical rationalist views of reflection—points made in the reply to Peacocke. I pay some attention to differences among the classical rationalists, especially ways in which Kant anticipates a more modern conception of reflection. I then take some small exploratory steps toward identifying positive aspects of reflection.

I consider three examples of what I consider to be successful reflection—one in meta-logic and two in philosophy. The example from meta-logic is the clarification of the notions of logical validity and logical consequence in the work of Skolem, Gödel, and Tarski.[28] The examples from philosophy are the attempts to clarify the error-presupposes-veridicality principle and the thought experiments that led to modern anti-individualism.

I try to isolate prominent features of these examples that yield clues to sorts of things reflection is especially well suited to understanding and clues to the ways

---

[26] See also the discussion of these matters in 'Postscript: Individualism and the Mental' in *Foundations of Mind*, *op. cit.*, 162–179.

[27] See also 'Frege on Knowing the Foundations', *Mind* 107 (1998), 305–347, reprinted in *Truth, Thought, Reason*, *op. cit.*

[28] A more detailed discussion of this case is set out in 'Logic and Analyticity' *Grazer Philosophische Studien*, 66 (2003), 199–249. I intend to include this paper in a later collection of papers.

in which reflection operates. All the cases center on a representational function of a cognitive activity, and on ways of fulfilling such functions. Two of the cases utilize insights into the form of a representational activity and into relations among representational abilities. All cases rely on judgments about necessary implications regarding the relevant subject matter. And all involve insight into explanatory priority. I believe that it is no coincidence that these features show up in the *meta-representational* enterprise of reflection. I think that reflection is better suited (though not uniquely well-suited) to be a distinctive and successful form of cognition when it focuses on representational powers than when it focuses on other matters.

The essay concludes by reviewing one of the primary oversights of the classical rationalist conception of reflection—its view that at least implicit understanding of the norms of reasoning is available to reflection for all reasoners. The classical idea was that the basic principles of reasoning are present, at least subliminally, in the psychologies of all reasoners. I believe that this view is clearly mistaken. I claim that this mistake applies even to critical reasoners. I reflect on what minimal resources must be present in any critical reasoner.

'Living Wages of *Sinn*' would have been included in the volume of my essays on Frege, *Truth, Thought, Reason: Essays on Frege*, if it had been written before that volume came out. I count the present volume as only second-best for inclusion of the essay, because the essay's explicit task is to discuss ways in which Frege's notion of *Sinn* figures in current philosophy of mind, philosophy of language, and epistemology. However, the main contribution of the essay, in my judgment, is its discussion of reflective understanding of representational contents. So the piece has some claim to be included here.

The essay begins by distinguishing Frege's notion of sense from modern notions of linguistic meaning. Senses are the ability-general aspects of thought contents that are associated, in certain specific ways, with linguistic expressions or linguistic uses.[29] Modern notions of linguistic meanings are contents of understanding linguistic expressions, contents that are associated with rules for using the expressions in idiolects or communal languages. The two notions, sense and linguistic meaning, are complementary theoretical tools for understanding language. They differ in their theoretical foci.

---

[29] Ability-general representational contents are those that individuate representational capacities (abilities) that are individuated independently of any particular, specific, occurrent exercise of it. Such representational abilities are freely repeatable in the sense that any exercise of the relevant ability need not be tied, anaphorically or in memory, to any other occurrent exercise of the ability in order for the ability to be the ability that it is. Conceptual and perceptual attributives are always ability-general. One could have the ability—either acquire or have it innately—in any number of occurrently different situations. For example, one could learn the concept piano in an isolated town in Russia or in New York City, in the nineteenth, twentieth, or twenty-first centuries. Ability-general representational contents contrast with occurrence-based *applications*—for example particular *occurrent uses* of a demonstrative or indexical word or representational content.

I defend this distinction, which I first articulated in 'Sinning Against Frege' (1979),[30] against criticisms by Saul Kripke. Here the issues primarily concern Frege's texts. Textual considerations help bring out how different Frege's conception is from modern conceptions.

I then criticize some of Frege's judgments about the senses of indexicals and proper names, using some of Kripke's important work on these matters. I maintain, however, that especially in understanding uses of demonstratives and in understanding the phenomenon of incomplete understanding, Frege's notion of sense is important to modern theorizing about language and thought.

The essay then focuses on Frege's seemingly disparate views on the individuation of senses. I argue that the main line of his thinking involves a very nuanced and sophisticated approach to understanding the form and content of psychological states, particularly propositional attitudes. The approach centers on recognizing the truth-conditional content and logical form involved in judgments and propositional inferences. I discuss various respects in which he notes that determining the content and structure of psychological states is a fallible, highly theoretical enterprise.

First, the sense associated with a linguistic utterance—for example, with demonstrative uses backed by perceptual beliefs—can vary significantly with context. There is no recipe for determining a contextually associated sense. Asking the individual what he or she was thinking is just a start to determining the nature of the thought. Second, the logical form of a sense can be determined only by considering many patterns of inferences that make use of it, as Frege's revolutionary method for determining truth-conditional logical form illustrated. Mere introspection does not suffice for understanding the truth-conditional structure of one's thought contents. Third, Frege's distinction between sense and coloring is valuable in separating truth-conditional aspects of a thought from psychologically relevant, collateral effects that are always associated with language usage. I argue that Frege was right not to allow coloring as a canonical, constitutive parameter in the structure of every thought. Commonly, coloring of one thought is the content of another thought. Applying the distinction in reflection is a theoretically delicate matter. Fourth, Frege's rationalist insights into *incomplete understanding* provide a further limit on reflective understanding of one's own thoughts, or those of others.

I have been asked, in a sceptical way, what difference is there between incomplete conceptual understanding and any other factually mistaken belief. I think that any defective conceptual understanding is factually defective. But not all factually mistaken beliefs constitute incomplete understanding, of a concept. I think it clear that some true beliefs that are expressible with a concept are not constitutive to understanding the concept. For example, mistakenly believing that

---

[30] 'Sinning Against Frege', *The Philosophical Review* 88 (1979), 398–432; reprinted in *Truth, Thought, Reason.*

there are thirteen rather than twelve pianos in a given region of space at a given time is not a defect in understanding the concept <u>piano</u>.

I believe that only radical philosophical ideology would maintain that incomplete understanding of a concept is *just any* sort of factual ignorance of a proposition containing it. One such ideology is an extreme extension of Quinean holism, according to which there is no distinction at all between change of belief and change of concept (as distinguished from no sharp distinction). Another such ideology is Leibniz's strange view that all facts (true propositions) expressible with a concept are embedded in the nature of the concept.

It is not part of my position that some, or any, concepts *are* embedded in a concept. My view on incomplete understanding is compatible with holding that all concepts are atomic. What I do believe is that connecting *some* (factual) propositions with a concept is constitutively relevant to understanding the concept, on an ordinary view of understanding. (There may be a family of such propositions, any group of which suffices.) Other propositions involving a concept are not constitutive to any level of understanding the concept.

At least with a non-defective concept—one that is not constitutively associated with a mistaken theory or a mistaken set of beliefs—the relevant propositions are true; they provide unifying generalizations about instances of correct applications of the concept; and they tend to justify and explain the unities—the fact that the instances are instances of the relevant concept.

I am not committed to thinking that, for finite minds, there is always such a thing as a complete mastery of the concept. Incomplete mastery could always be relative to a better, more deeply explanatory explication. Or the notion of complete mastery can be too vague to be fruitful. I am also not committed to believing that there are *sharp* boundaries marking any of the distinctions among levels of understanding. Moreover, I think that in most cases, including both empirical and mathematical cases, no explication, no matter how good, *fixes* the *representatum* or extension of the explicated concepts. Nearly always, the *representatum* is fixed by attributional applications to cases, together with objective similarity relations among cases.

I want to highlight two values to taking incomplete understanding seriously as an important idea in philosophy. One is that doing so underlines the important fact that having a constitutive explication (a general criterial understanding) is almost never necessary for *having* and using a concept. One can have a concept without fully understanding it. Having concepts is grounded in applications to cases, and in not fully articulated or generalized senses of patterns of similarities among cases. That is why, in philosophical understanding, examples tend to drive progress in reflection. The second value of taking incomplete understanding seriously is that it combats the deeply entrenched but deeply mistaken idea that we have a transparent grip on—understanding of—the nature of our own concepts. The deepest progress in factual knowledge is often concomitantly a progress in conceptual understanding.

In 'Living Wages of *Sinn*', I distinguish different types of incomplete under-standing and discuss limitations on reflection. I have discussed such limitations in other work. I do so more fully and systematically here. By reference to the post-Newtonian elaboration and refinement of Newton's concepts in the calculus, I provide a relatively detailed account of issues that complicate our understanding of reflective conceptual explications. I highlight interplay between reflection and new mathematical work. And I further develop criticism of the traditional conception of reflection. According to this conception, reflection brings to con-sciousness conceptual explications that are unconscious but already fully formed, or at least fully available in the unconscious psychology of the reflective thinker, in the sense that all elements for forming the explication are present in the psychology.

This collection invites the reader to make connections and recognize systematic structures that were not fully in mind when the essays were written. I have sketched some of these connections and structures in this introduction. I think that in the best philosophical work, a philosophical unconscious guides the author's thinking in incompletely understood ways, ways that point fruitfully beyond what is realized, consciously or even unconsciously, at the time the work is produced. This phenomenon is a natural consequence of the way in which thought is molded by a subject matter that is larger and more complex than can be fully grasped at any given time, or indeed, sometimes, ever. I hope that, in reading these essays together, some readers will have the pleasure of thinking constructively, not just critically, beyond where any given essay manages to reach. I hope myself to continue to be one such reader.