7 *Self and Self-Understanding**

Lecture I: Some Origins of Self

I will reflect on constitutive features of selves—especially a certain sort of self-understanding. This self-understanding is the main topic of these lectures.

I

'Self' is a technical term, refined from ordinary usage. Ordinary usage is, however, very close to what I want. A definition from the *Oxford English Dictionary* runs, 'Self: a person's essential being that distinguishes the person from others, especially considered as the object of introspection or reflexive action; a person's particular nature.' Kant characterized a *person* as 'what is conscious of its numerical identity, of its self, in different times'.[1] This

* This essay is a revision, with some expansion, of the Dewey Lectures, given at Columbia University, December 2007. I am grateful to Christopher Peacocke for valuable criticisms in spring 2011 of the last section of Lecture I and all of Lecture III; and to Denis Bühler for saving me from an error and prompting an argument in Lecture II.

[1] Immanuel Kant, *Critique of Pure Reason*, A361; see also *Metaphysics Mrongovius* 29: 911 in Karl Ameriks and Steve Naragon (eds.), *Lectures on Metaphysics* (New York: Cambridge University Press, 1997), p. 276. Kant's formulation in the *Critique* is not ideally specific and might qualify as ambiguous. It is clear from context that Kant means the consciousness to include consciousness at a given time of the self as it is at different times. It is also clear that Kant intends the consciousness to be noninferential, and in my terms *de re*. Kant shows less interest in the *diachronic* implications of his formulation than one might hope. He relies on these implications in the Third Paralogism to argue that cognition of self over time is necessarily empirical and cannot meet requirements of rationalist theories of self. In my 'Memory and Persons', *The Philosophical Review* 112: 3 (July 2003), 289–337, note 50, I sketch why I reject Kant's argument that self-cognition of self over time is necessarily empirical. In notes from his lectures, some characterizations of *person* or *personality* omit reference to transtemporal self-consciousness altogether, citing only a being conscious of its identity in different states. See *Metaphysics* L_1 28: 276–277; *Metaphysics Dohna* 28: 680. Both passages are in Ameriks and Naragon (eds.), *Lectures on Metaphysics*, 87–88, 381. See also *Critique of Practical Reason* 5: 87, in Mary J. Gregor (ed), *Practical Philosophy* (New York: Cambridge University Press, 1996), 210. Kant thought that self-consciousness necessarily involves inner sense. There is ample evidence that he assumed that inner sense necessarily has past and future as well as present applications. So these failures to mention transtemporal capacities do not reflect a different position from the one in *Critique of Pure Reason*. But they do show that he does little to develop this aspect of the notion. In *The Metaphysics of Morals* 6: 223—in Gregor (ed.), *Practical Philosophy*, 378—Kant

conception of *person* is very close to the notion of a *self* that interests me. It is a purely psychological notion that specifies a self-consciousness with diachronic reach. Regardless of the ontology of selves—whether or not they are purely psychological beings—the *concept* self is a psychological concept.

Kant's concept contrasts with Strawson's concept person. Strawson's concept is close to the common-sense notion of person—roughly, the sort of bodily being—paradigmatically human being—that normally matures into a self-conscious critical reasoner with moral capabilities. Strawson's *concept* is *not* purely a psychological concept. It *entails* that persons have physical as well as psychological characteristics. These two notions—the OED/Kant notion and Strawson's notion—typify a distinction that I draw between selves and persons.

Selves are *loci*, indeed agents, of psychological activities, and agents that can engage in, and are subject to, certain valuations—the distinctive valuations associated with persons. Selves evaluate themselves and other selves for rationality, critical rationality, morality, social cooperation, character, creativity, grace. Being an agent and topic of such valuations is part of what makes them selves. Since selves constitute the particular psychological natures of persons, selves' being agents and topics of such valuations is part of what makes persons *persons*.

In engaging in these types of valuation, selves become the subject matter of self-knowledge and self-understanding. So the valuational and cognitive powers of selves are intertwined. Some of the types of evaluation that I listed are constitutive to being a self. Selves and persons are constitutively capable of evaluating, and engaging in, critical reason. Arguably, they are constitutively capable of being moral. A certain reflexive cognition is constitutive to these evaluations. Critical reason and morality constitutively depend on self-understanding. My topic is the psychology and epistemology of the kind of self-understanding that is required for critical reason and morality, and that is itself constitutive to being a self, and a person.

First, some introductory points. I am interested in *constitutive* matters—in the *nature* of selves—those features of selves that make them selves.[2] I characterized selves as certain *loci* of *psychological* realities. The fact that self is a purely psychological notion does not entail, or even much encourage, the view that selves lack physical natures. The psychological nature of the *concept* cannot legislate these ontological matters. I will, however, be focusing on the *psychological* nature of selves. I bracket any further, non-psychological aspects of their nature.

distinguishes moral personality—a free being with reason—from psychological personality, explained in terms of consciousness of self in different states. (See also *Metaphysics* L_1 28: 276–277) The two notions are connected, however, inasmuch as a free being with reason must be capable of remembering and anticipating acts if acts are to be imputed to the individual's moral personality.

[2] I discuss natures in my *Foundations of Mind: Philosophical Essays, Volume II* (New York: Oxford University Press, 2007), 1–3 and *passim*; and in *Origins of Objectivity* (Oxford: Oxford University Press, 2010), 57–67 and *passim*. For working through *passims*, see the indexes.

It is natural and right to connect psychological discussion with discussion of the brain. I will have almost nothing to say about the brain. I think that a psychological framework is necessary to finding anything in the brain that is of psychological interest. Ultimately, psychology and neuroscience are collaborative enterprises. Each must provide checks and balances on the other. But a psychological framework will provide us with more than enough material to work with here. The issues about understanding and value that I will be discussing *must* be connected to psychological inquiry. We cannot understand these issues in non-psychological terms. Perhaps this situation will change. But I doubt it.

The tradition of discussing selves is complex. There are several closely related concepts that figure in the tradition—soul, spirit, mind, subject, conscious subject, self-conscious subject, rational animal, rational being, critically rational being, ego, person. Several of these will appear in my discussion. First, I will return to contrasting my notion of self with that of person.

Strawson's elaboration of the notion of person has strongly influenced philosophy in the last half-century. I think that this influence is deserved. But it has led to misconceptions about relations between psychological and physical attributions. I will not have time for detailed discussion of Strawson's views. I will just describe them and state some attitudes toward them.

Strawson developed his concept of persons to combat dualist conceptions of mind.[3] According to his concept, a person essentially has both psychological and corporeal attributes. Strawson's key claim is that use of the concept person, with its essential application to a corporeal being, is necessary for understanding psychological concepts. In other words, application of psychological concepts would be incoherent and unintelligible if they were not applied to something, a person, that has corporeal as well as psychological attributes.

I believe that none of Strawson's arguments, or any by his followers, justifies this claim.[4] All depend on views about knowledge or individuation of content that are neither plausible nor well supported. I do think that Strawson's notion person is useful. What I reject is a set of arguments for holding that all psychological notions are conceptually dependent on a prior notion of a corporeal being, a person. I will take psychological notions on their own terms. I emphasize that these are conceptual, constitutive points. I will not be defending an ontological position.

It is only because mature persons naturally have certain psychological capacities that we evaluate them for critical rationality, take them to have a special

[3] P. F. Strawson, *Individuals* (1959) (London: Routledge, 2002), chapter 3.

[4] Strawson argues for his views in *Individuals*. See also his *The Bounds of Sense* (1966) (London: Routlege, 2006), 163ff. For followers, see Gareth Evans, *The Varieties of Reference* (New York: Oxford University Press, 1982), 208ff., 237ff.; John McDowell, *Mind and World* (Cambridge, Mass.: Harvard University Press, 1994), 100ff.; John Campbell, *Past, Space, and Self* (Cambridge, Mass.: MIT, 1995), 92ff.; Quassim Cassam, *Self and World* (New York: Oxford University Press, 1997), passim; and most of the essays in José L. Bermúdez, Anthony Marcel, and Naomi Eilan (eds.), *The Body and the Self* (Cambridge, Mass.: MIT, 1995).

moral status, and so on. Not all persons have these states and capacities at all times. Six-month-olds are persons. But if a being naturally and constitutively did not have or develop psychological states and capacities beyond those *exhibited* by six-month-old human beings, they would not be persons.[5]

Perhaps persons *develop* into having or being selves. They would do so by developing the competencies that meet certain conditions for falling under norms constitutive of being selves, such as norms for engaging in critical reasoning or for being morally responsible. Then persons would not have or be selves during all times when they, the persons, exist. Demented human individuals that lack psychological capacities beyond chimps are persons, but only because their natural capacities have been distorted by age, damage, or disease. We understand our concept person partly by reference to certain natural psychological competencies. The reference is to naturally having the competencies in a mature state, not to having them throughout a person's existence. I think of selves as beings that actually have the competencies that make persons the valuable beings that they are. Thus selves are types that set standards for being a person. The standards allow for prolepsis and retrospection.

It is compatible with the main part of my discussion to hold that mature selves are just phases in the lives of persons. There are intuitions that pull in that direction. On the other hand, I find it natural to reify—to say that someone has gone out of existence when a person still exists, but has irrevocably lost those psychological capacities that made the person a person. I am inclined to count selves as individuals. I think that natural instincts and even explanatory and normative considerations suggest being liberal in allowing existence to different types of being, even where they overlap in seemingly messy ways. Here, these issues will not matter. I focus on constitution, nature, essence. I shall speak of selves, for the sake of exposition, as individuals.

Like most psychological notions, the notion of self bears complex constitutive relations to a wider physical reality. As far as I know, selves may necessarily depend on a physical body. I think, however, that there is no evident *conceptual* or even apriori necessity that selves have physical properties, although all selves that we know of depend in some way on their bodies.

Selves are necessarily conscious, at least some of the time. A type of consciousness presupposed by all other types is phenomenal consciousness.[6] Phenomenal consciousness consists in there being some way that it is like to be in a mental state. Phenomenal consciousness is an occurrent condition. It is the psychological bedrock of selfhood. An individual's stream of consciousness is

[5] I write of *naturally* having competencies, or of *natural* competencies, to allow that a self or person could be damaged or diseased so that performance that would realize the competencies is blocked. If a self permanently loses such competencies, it no longer exists or the person no longer is a self.

[6] The notion of phenomenological consciousness is vividly illustrated in Thomas Nagel, 'What Is It Like to Be a Bat?' *The Philosophical Review* 83:4 (October 1974), 435–450.

interrupted, in sleep for example. But phenomenally conscious psychological life is where the acts, cognitions, and evaluations that most matter to us occur.[7]

Underlying conscious life is, of course, a vast system of unconscious psychological states and activities. In empirical explanation, unconscious elements are primary. But unconscious psychological states and occurrences are constitutively relevant to selves only insofar as they bear certain relations to conscious states and occurrences.

I will concentrate here on the *representational* nature of selves. Selves are subjects with representational competencies, states, events, and acts. They are subjects with perceptual systems, capacities for inference, beliefs, intentions, perceptions, occurrent thoughts, decisions.

Now I introduce a technical term. The *point of view* of an individual is the system of representational states and occurrences that are, in a certain sense, correctly imputable to the individual.[8] Non-human animals as well as persons and selves have points of view. Imputability is a partly intuitive, partly theoretical notion. I start with intuitive points. *Non-psychological* entities are not part of a point of view. Irretrievably unconscious modular *psychological* states and processes are also not part of a point of view. They are not imputable to the individual, though they occur in the individual's psychological subsystems. Potentially or actually conscious perceptions, memories, beliefs, intentions, emotions, decisions, imaginings are part of a point of view. So are attitudes in the Freudian unconscious.

I believe that the key to the nature of selves lies in reflecting on powers distinctive to their points of view. These powers include competence for a certain type of reflexive self-attribution of elements in the point of view. Such self-attributions give the type of point of view that is distinctive of selves a *multi-tiered* structure. I shall explore this structure. It is the structure of self-consciousness and representation of self.

[7] Compare William James, *The Principles of Psychology*, vol. 1 (Cambridge, Mass.: Harvard University Press, 1981 [1890]), chapter 10, 284–285, which identifies the self with '. . . either the entire stream of our personal consciousness, or the present "segment" or "section" of that stream, according as whether we take a broader or a narrower view.' He calls the narrower, more present segment the 'innermost centre within the circle, or sanctuary within the citadel, constituted by the subjective life as a whole. Compared with this element of the stream, the other parts, even of the subjective life, seem transient external possessions, of which each in turn can be disowned, whilst that which disowns them remains. . . . It is the home of interest. . . . It is the source of effort and attention, and the place from which appear to emanate the fiats of the will.'

[8] By extension, I count the representational contents and non-representational conscious features of the relevant psychological states the individual's point of view. I emphasize that '*point of view*' and '*perspective*', which I use interchangeably, are *not* to be construed as indicating anything essentially perceptual or empirical. One can have a purely mathematical point of view or perspective. Points of view and perspectives are ways of representing (thinking about, perceiving, remembering) a subject matter. I confine these terms to "ways," and states marked by those ways, that are imputable to an individual, rather than subsystems of individuals.

II

I begin by discussing some background conditions and some representational antecedents of the representation of self that is distinctive of selves and persons.

Every living system can differentiate itself from other things. It expels foreign matter as waste, ingests other bodies and not its own, protects itself from external threats, and so on. These capacities are not in themselves either representational or psychological. They form the functional background for self-interest and representation of self.

Perception is the lower border of sensory representation.[9] Not all sensory systems are perceptual. Hence not all are representational. Sensings of heat or light by amoebae or worms and the sensors in muscle tissue that signal stress are not representational in my sense. The explanatory paradigms that drive the theory of perception differ fundamentally from those that drive explanations of the sensitivities in these sensory systems. Explanation of perception makes essential reference to veridicality conditions. Explanation centers on how the system represents the surrounding environment accurately, to the extent that it does. Of course, even the simplest sensory systems exhibit functional success and failure. A sensory system is functionally successful when its response to stimulus is associated in the environment with the macro-qualities that it evolved to respond to. But in non-perceptual sensory systems, success is only functional good fortune. There is no explanatory value in invoking veridicality conditions.

By contrast, perceptual systems are explained in terms of the formation of states that are about the distal environment and that are veridical or not. The mark of a perceptual system is *perceptual constancy*. Perceptual constancies are abilities systematically to represent given distal features, by sensory means, *as* those same features, even though the proximal stimulations caused by those features vary radically. For example, most perceptual visual systems can represent something as having a given size as it moves closer or farther away—thus as it causes radically different proximal stimulations. Perceptual constancies occur in the visual, auditory, and touch systems of a wide variety of animals, including the visual systems of certain insects.

As noted, all living organisms show self-interest, broadly understood. I am interested in antecedents of *representation* of self. The most primitive representational antecedent is constitutive to all perceptual and actional representational content. This antecedent is an *ego-centric index*.[10] In their most primitive forms,

[9] I defend this claim and the associated conception of representation in *Origins of Objectivity*, especially chapters 8 and 9. A state is representational in my sense if and only if specification of the state in terms of veridicality conditions could ground serious, correct explanation.

[10] Another term for such representation is '*de se* index'. For discussion of such indexes, see 'Memory and Persons', and *Origins of Objectivity*, 199–201, 287–88, and passim. For other discussions, see Marc Jeannerod, *The Cognitive Neuroscience of Action* (Cambridge, Mass.: Blackwell, 1997); Jeannerod, 'To Act or Not to Act: Perspectives on the Representation of Actions', *Quarterly Journal of Experimental Psychology* 52A:1 (1999), 1–29; J. D. Crawford,

such indexes are not conceptual. Applications of ego-centric indexes anchor all perception and all primitive actional states. In primitive forms, they are not used or applied *by the individual*. They are, however, a primitive type of ego-related representation. These primitive forms become incorporated into propositional perceptual beliefs and propositional intentions. Then they are conceptual.

In perception, ego-centric indexes refer to the individual—or, more commonly, some position on the individual or some time at which the perception occurs—as an *origin* with respect to which representation of other entities is mapped. Ego-centric indexes single out an individual, position, or time in a context-dependent way. For example, an index might mark a spatial origin in relation to which other entities are spatially related in a perception. A nest-perception might represent the nest as to the left at such and such a distance from the origin of the perception. The origin is perhaps a position between the eyes of an insect. Or a perceptual memory of a food source might represent it as having been visited at some time interval from the time represented by an ego-centric index as the present time of the memory representation.

Like the self-concept, ego-centric indexes are immune to reference failure. They automatically mark an origin of reference that occurs whenever a state with the relevant index is instantiated. Unlike applications of the first-person concept, occurrent applications of ego-centric indexes are normally not acts. Applications depend purely on their role in a larger system of representation. Unlike applications of the first-person concept, not all ego-centric indexes *refer* to an individual or ego. Many refer only to a position or a time.

Ego-centric indexes have two constitutive functions. One is the function just discussed—to index an origin for a framework of representation. The other is to type-individuate an aspect of psychological states that treats the origin as having immediate ego-relevance. All representational psychologies function to privilege the individual's needs, goals, and perspective, and function to provide the representational basis for serving the individual's needs, goals, and perspective—doing things for the individual from the individual's own motivations or other powers. The applications that realize the first of the two constitutive functions of ego-centric indexes—applications to spatial or temporal positions—are always themselves privileged in an individual's psychology. They function as privileged in that they function in the psychology to have psychologically immediate ego-relevance. They mark spatial origins of perception and spatial positions from which the individual's actions are initiated. They mark positions that the individual must protect in serving the individual's own needs, goals, and perspective. They mark times, present times, for initiating action and calibrating memories and anticipations. They mark times that are privileged with respect to serving the individual's needs, goals, and perspective.

W. P. Medendorp, and J. J. Marotta, 'Spatial Transformations for Eye–Hand Coordination', *Journal of Neurophysiology* 92 (2004), 10–19.

Like all psychological representational content, the ego-relevant content of spatial and temporal ego-centric indexes marks representational functions that are in turn grounded in representational competencies. The ego-relevant content marks competencies to treat indexically anchored positions as privileged in realizing functions to distinguish and serve the individual's own needs, goals, and perspective. All representational psychologies have these functions and competencies. All mark temporal or spatial positions indexically in an ego-relevant way. The competencies that underlie these primitive ego-related representations are enriched in individuals that have complex psychologies, until they are part of the suite of competencies that are constitutive of persons and selves. Ego-centric indexes are the ur-antecedents of concepts of self.

Of course, an insect cannot spell out this ego-relevance. It has no attributives that apply to psychological states or content. Ego-centric indexes, however, are part of the representational content of—and help individuate—psychological states that systematically relate a framework origin to the individual's needs or perspective, non-inferentially and as immediately as the system allows. This functional relation to the individual's needs, goals, or perspective is what makes an index—and the psychological state that it helps individuate—ego-centric.

The relation to the individual's needs, goals, or perspective usually runs through capacities mediating perception and action. For example, the psychological system might be such that if an object is perceived to be approaching the origin, the animal will take protective measures. Or certain feeding actions might be initiated if a memory marks a food source as having been visited at a certain temporal interval from the present.

Ego-centric indexes and their two functions are constitutive to every representational psychology. Every representational psychology contains some states that include them. Powers to realize these functions help mark an individual with a representational mind as an individual.

Fulfilling these functions involves a representational capacity to distinguish and privilege the individual's own needs, goals, and perspective. By distinguishing and privileging the individual's needs, goals, and perspective, ego-centric indexes provide a representational basis for serving them. The unity of an individual's representational psychology consists partly in having representational powers that mark off and unify that psychology by fulfilling the individual's needs and goals, and otherwise serving its representational perspective.

As noted, ego-centric indexes are ancestors of the first-person concept, expressed by the word 'I', that represents persons and selves. The indexes' infallible reference, their anchoring a representational framework, and their direct ego-relevance all presage aspects of the first-person concept. Having ego-centric indexes does not require a capacity to *represent* anything *as* having psychological states. In this respect, they differ from a full concept of self.

III

In the remainder of this lecture, I discuss four capacities that mark milestones in the developmental prehistory of distinctively *self*-representation. The first three are more complex exercises of ego-centric indexes. The exercises are psychological couplings of ego-centric indexes between and within sensory modalities. These couplings are, I think, early precursors of a certain capacity for objectification that is characteristic of selves and persons. I have in mind the capacity to think of oneself from first- and third-person points of view. In their most primitive forms, these couplings occur among sensory capacities. They are non-conceptual. Non-human animals exhibit at least the first capacity, and possibly all three. They are thus nodes in the "animal" roots of selves. The fourth capacity is a type of memory that yields another kind of objectification constitutive to selves and persons.

Let us consider the first of these four capacities. In the last three decades, ethology and developmental psychology have been astir over beginnings of self-awareness. Some of this discussion is onto important junctures in the prehistory of personhood and selfhood. The literature has, however, persistently blurred important psychological distinctions. Much of it vastly overstates the directness of connection between the phenomena being theorized about and self-awareness in any sense that involves representation of a representer's own psychology.

I start with what is called '*the mirror test*'. A mark is applied to a bodily surface without an individual's being aware of the application. The mark is placed so that the individual can see it only in a mirror. An individual passes the mirror test if after a short time, it acts as if its own behavior is the source of the behavior shown in the mirror.[11] For example, the individual might touch the mark. Chimpanzees, elephants, and dolphins pass the test. Only a few gorillas in captivity have passed it. It was long thought that although monkeys can learn to use mirrors to spot other objects, they cannot pass the test. Recently, there is some evidence that rhesus macaques pass it. Human children first pass it between 16 and 24 months.

Passing the test has been taken as evidence for self-awareness.[12] The relevant notions of *self* and *awareness* need clarification. There is no clear sign that in passing the test an animal represents itself as having psychological states. The

[11] Gordon G. Gallup, Jr., 'Chimpanzees: Self-Recognition', *Science* 167: 3914 (Jan. 2, 1970), 341–343; Gallup, 'Self Recognition in Primates: A Comparative Approach to the Bidirectional Properties of Consciousness', *American Psychologist* 32:5 (May 1977), 329–338; Sue T. Parker, Robert W. Mitchell, and Maria L. Boccia (eds.), *Self-Awareness in Animals and Humans: Developmental Perspectives* (New York: Cambridge University Press, 1994); Diana Reiss and Lori Marino, 'Mirror Self-Recognition in the Bottlenose Dolphin: A Case of Cognitive Convergence', *Proceedings of the National Academy of Sciences* 98:10 (May 8, 2001), 5937–5942.

[12] Gallup, 'Self-Recognition: Research Strategies and Experimental Design', in Parker, Mitchell, and Boccia (eds.). *Self-Awareness in Hun and Animals*. Gallup, the originator of the test, gives the following experimental condition on exhibition of self-awareness: 'An organism is self-aware to the extent that it can be shown capable of becoming the object of his own attention.'

experiments do not indicate possession of what I count a self-concept or first-person concept. Still, there are steps in those directions.

Passing the mirror test rests on bidirectional couplings in the individual's psychology between visual perceptions and representations in kinesthetic proprioception. The individual can match what it sees with what it senses in its own movements. The individual's psychology translates between visual representations and kinesthetic/proprioceptive representations as of the same movements—tokens and types. The translation is systematic. The mapping seems to use a continuing image of the whole body, built through proprioception in the individual's learning history. Mirrors are not essential for such couplings, but passing the mirror test evinces them.

The systematic character of the couplings and the fact that they operate on a whole-body image differentiate the behavior exhibited in passing the mirror test from the behavior of feeling an irritant on a body part, finding it visually, and brushing it off. This latter type of behavior also involves couplings between different sense modalities. It does not, however, suggest visual identification of something, codified in the body schema, as the whole individual.[13] Visually locating and dealing with an irritant does not depend on systematic whole-body coordination between vision and proprioception. Visually identifying as such what has been mapped proprioceptively *as of a whole individual* is a kind of reflexive "self" identification. Here the "self" is the whole individual. The individual is identified in terms of bodily characteristics. The identification is, as far as the experiments show, at the level of intermodal connections between perceptual capacities, not necessarily effected through propositional thought.

The connection between vision and *kinesthetic* representation makes possible not only the individual's representation of its whole body, but makes possible a kind of representation of the body *as the individual's own*—that is, something like an *attribution* of *ego-indexing* to the individual's own bodily actions. Let me explain this point step by step.

First, ordinary visual representational content is ego-centrically indexed. But such content *attributes* to *seen* particulars only physical or functional characteristics (shape, position, color, danger, food, and so on). Primate vision represents body parts and bodily movements as such. It probably also differentiates purposive from non-purposive movements.[14]

[13] For scientific discussion of the notion of body schema, see J. Paillard, 'Body Schema and Body Image: A Double Dissociation in Deafferented Patients', in Gantcho N. Gantchev, Shigemi Mori, and Jean Massion (eds.), *Motor Control: Today and Tomorrow* (Sofia, Bulgaria: Academic Publishing House, 1999), 197–214. For philosophical discussion, see Brian O'Shaughnessy, 'Proprioception and the Body Image', in Bermúdez, Marcel, and Eilan (eds.), *The Body and the Self*, 175–204.

[14] B. Hare et al., 'Chimpanzees Know What Conspecifics Do and Do Not See', *Animal Behaviour* 59:4 (April 2000), 771–786; Laurie R. Santos, Jonathan I. Flombaum, and Webb Phillips, 'The Evolution of Human Mindreading: How Nonhuman Primates Can Inform Social Cognitive Neuroscience', in Steven M. Platek, Julian Paul Keenan, and Todd K. Shackelford (eds.). *Evolutionary Cognitive Neuroscience* (Cambridge, Mass.: MIT, 2007), 433–456. A lot of theory in

Second, the bodily movement seen in the mirror causes not only visual representation but a kinesthetic sense of the movement. The representational content of kinesthetic perception, like that of visual perception, is ego-centrically indexed—marking the origin of the perception's perspective. But there is an important difference between proprioception and vision. In proprioception, ego-centric indexes also tag the *perceived* entities. Proprioception marks the perceived bodily movement as having the same relevance to the individual's needs and perspective as the proprioceptive perceptions themselves have. A function of proprioceptive perception is to track the animal's movements *as* the animal's own. So whereas, by the point of the previous paragraph, ordinary visual representations are singly indexed, kinesthetic perceptual representations are always *doubly* indexed. In addition to the index, or indexes, for the perspectival origin of the perception, kinesthetic perceptual content associates an ego-centric index with the *perceived* particulars. Proprioceptively perceived bodily movements are marked as privileged with respect to the individual's needs, goals, and perspective.

Third, the proprioceptively perceived body parts and movements are mapped onto a standing proprioceptive body schema. Like all proprioception, this schema is doubly indexed: both representational origin and proprioceptively perceived entities are indexed.

Fourth, for individuals who pass the mirror test, mirror experiences set up an intra-psychological coupling between visual representations of bodily movements and kinesthetic action representations of the same movements. These couplings induce a further coupling between the visual representations of body parts and positions on the body schema.

Fifth, these couplings provide a causal pattern that grounds new representational content. There is a new association of ego-centric indexes with *objects* of *visual* perception. Thus *seen* objects—say, the movement of a body part—are directly and systematically coordinated in the psychology with the individual's needs, goals, and perspective. The movement is seen as ego-privileged. Single-indexed and double-indexed visual perceptions represent the same movement. But the latter are constitutively linked to kinesthetic and body-image representational contents.[15]

this domain over-attributes representation of psychological states to apes and very young children, when attributing representation of teleological states suffices to account for the evidence.

[15] The association of single-indexed and double-indexed visual representations is informative. Such associations can in principle be mistaken. The usual visual representation of the form

$[ego_i]\ that_1$ movement

comes to be associated with a visual representation of the form

$[ego_i]\ that_1\ [ego_i]$ movement.

Both represent the same movement. At the level of a psychology that passes the mirror test, the ego-centric indexes may well refer to the whole body rather than merely a body part.

Earlier I said that ego-centric indexes are infallibly referentially successful in their primary function. In associating ego-centric indexes with perceived entities, a perceptual system is fallible

The application in vision of single-indexed and double-indexed representational contents to the same movement is not a propositional identity. It is a coordination between perceptual representations in the psychology. But the visual system functions to treat the different representations as representing the same movement.[16]

Passing the mirror test marks three developmental milestones. They may have been achieved before the test is passed. Passing the test just shows that they have been achieved.

First, by associating an ego-centric index with a purportedly visually perceived physical entity, the psychology invests that entity with the same ego-significance that the origin of the visual perception has. Recall that the second function of ego-centric indexes is to mark a direct psychological relation between the individual's needs, goals, or perspective and what they index. What is new is that this indexing is associated with a *visually* perceived entity. No more in this new step than in the most primitive ego-centric indexing does the individual use psychological attributives that might figure in a *theory* of mind. No psychological states are represented (referred to or indicated) by ego-centric indexes.

Second, the individual's psychology associates occurrences of its ego-centric indexes with the individual's whole, visually perceived body, not just with an indexed anchor *position*. As noted, the proprioceptive body schema constitutively marks a perceptually represented whole body, and does so in a doubly indexed way. It is not part of the *nature* of the visual system to doubly index anything. In coming to associate an ego-centric index with a visually perceived body, the psychological system has an intra-psychological functional connection between the anchor of the visual perceptual representational content and the perceived body. (Of course, perceived body parts can also be associated with an ego-centric index and mapped as parts of the doubly indexed whole body.) As with the first milestone, what is new lies in the linking of visual competence with double-indexing that is already present in another system.

A third step is implicit in the second, and is, I think, the most significant one. Passing the mirror test evinces a step toward *objectification* of self in two respects. One is intermodal. The development of connections among kinesthetic proprioception, the proprioceptive body schema, and visual perception allows the perceiving individual to associate ego-centric indexes with the same individual as

inasmuch as perceptual representation of perceived entities can fail. So association of an ego-centric index with a purportedly perceived entity can participate in a failure to connect with purportedly perceived reality. However, even such an index marks a relation in the psychology of an origin in perception to the individual's needs, goals, and perspective. Occurrences of an ego-centric index cannot fail in representing the origin or in connecting a purported perceptual attribution to the individual's needs, goals, and perspective.

[16] This association ramifies the perceptual constancies that mark all perception. I conjecture that a visual system cannot produce doubly indexed representational contents on its own. A doubly indexed visual representation always depends on some intra-psychological connection to a naturally doubly indexed representation in another perceptual modality—specifically, proprioception—or in a conceptual application. In proprioception, the individual perceives his or her own bodily arrangements or movements, and naturally perceives them as being ego-related.

perceived from the perspectives of different perceptual modalities. The perceived individual is perceived as having ego-significance, in this limited way, from different perspectives; and the perspectives are linked with one another and with actional representations.

The other aspect of objectification concerns visual representation. Couplings among vision, kinesthetic sense, and body-schema representation make possible an informative connection of single-indexed visual perceptions with double-indexed visual perceptions of the same entities. A visual perception of a face—anchored as usual by an ego-centric index—may represent a face as just another face. Then when the intermodal connections are established, the singly indexed perception comes to be informatively associated with a visual perception that attributes an ego-centric index to the same face. So the second perception is double-indexed. The face is seen from two perspectives. (Proprioception is always double-indexed and does not provide two perspectives.) One perspective characterizes one's face, but not as ego-related. The other characterizes the identically appearing face as the individual's own. The case is a primitive, preconceptual ancestor of "Aha, that's me!" It is a primitive ancestor of representing oneself from both first-person and third-person perspectives.

To describe these advances as an emergence of *self*-consciousness or *self*-representation would be misleading. Ego-centric indexes are associated with an object of representation. But no psychological states are attributed to it. So no individual is represented *as* a self. Reflexivity in representation should not be confused with genuine *self*-consciousness or *self*-representation. Still, the reflexivity and coupling of different points of view shown in the mirror test is a *step* toward genuine self-representation.

IV

The psychology evinced in the mirror test is a simple case of intermodal and intramodal perceptual couplings of ego-centric indexes. Such couplings occur in more complex forms. Each form is of interest in itself. Each adds a distinct type of objectification in ego-centric indexing. I will discuss two other couplings—*imitation* and *joint attention*.

Imitation starts from the beginning of human life.[17] Children imitate the expressions of parents from the first few minutes after birth. Purposive social

[17] Andrew N. Meltzoff and M. Keith Moore, 'Imitation of Facial and Manual Gestures by Human Neonates', *Science* 198:4312 (Oct. 7, 1977), 75–78; Alison Gopnik and Meltzoff, 'Minds, Bodies, and Persons: Young Children's Understanding of the Self and Others as Reflected in Imitation and Theory of Mind Research', in Parker, Mitchell, and Boccia, (eds.), *Self-Awareness in Humans and Animals*, 166–186. This copying may be associated with mirror neurons—neurons activated both when an individual observes specific types of activity and when the individual performs the same specific types of activity. Mirror neurons were discovered by Vittorio Gallese et al., 'Action Recognition in the Premotor Cortex', *Brain* 119:2 (April 1996), 593–609. Mirror neurons are popularly construed as

copying plays a major role in the emergence of more refined systematic couplings among an individual's sensory modalities and, of course, in the development of language. *Copying* is matching in which one specific type of behavior in one individual is caused by, and specifically explained by, the sensing—perceptual or not—of the same type of behavior in another individual. Copying divides into several interestingly different subspecies, of which imitation is the most important for the development of a self-concept. *Imitation* is active, purposeful, goal-directed copying of a specific form of behavior. Examples of non-imitative forms of copying are *mimicry*—an automatic sensory-motor type of copying—and *emulation*—purposive copying whose point is to match a behavioral *result*, not necessarily the behavior itself. It is disputed whether non-human animals ever imitate as opposed to emulate.[18] Non-human animals are certainly more oriented toward results than toward coordinating behaviors themselves. From 12 months onward, human children differ dramatically from other primates through greater orientation to imitation, and less to emulation.

Imitation again involves systematic coupling among sensory modalities, and between sensory modalities and the representational actional system. Suppose that the imitating individual has a visual perception, (V)$\underline{[\text{ego}_i]\text{that}_1\text{F}}$. ('(V)' designates the visual modality. '$\underline{\text{F}}$' indicates an attributive mode of presentation that is perceptually applied to the token behavior *A* of the imitated individual. I assume that the ego-centric index is spatial. I ignore the inevitable temporal index. The subscripts on 'ego' and 'that' mark occurrent applications of the index and the demonstrative-like capacity, respectively.) The individual adjusts his action—and his psychology adjusts representation of action—so that it is of the same *type* as the imitated action. (V)$\underline{[\text{ego}_i]\text{that}_1\text{F}}$, referring to *A*, is taken as the paradigm model for the individual's imitation. The imitating individual produces a type-similar imitating action A_i in such a way that it satisfies a representation $\underline{\text{F}}$ as of the same type of action as the imitated action *A*. So the psychology produces in its actional system and, presumably in its kinesthetic proprioceptive system, doubly indexed representations (Act)$\underline{[\text{ego}_i]\text{that}_2[\text{ego}_i]\text{F}}$ and (K)$\underline{[\text{ego}_i]\text{that}_2[\text{ego}_i]\ \text{F}}$. ('Act' indicates actional representational mode; 'K' stands for kinesthetic mode.) The actional representation anticipates the imitating act A_i. The kinesthetic representation is caused by A_i. Given that the imitation is successful, acts *A* and A_i are in fact of type F. Imitation is a purposive coupling between the agent's own act A_i and another individual's type-similar act *A*. The coupling of the visual representation (V)$\underline{[\text{ego}_i]\text{that}_1\text{F}}$, which perceptually represents the act *A* that is to be imitated, with

instantiating "theory of mind" or "mind reading". This construal over-intellectualizes the phenomenon. The function of mirror neurons has not been agreed upon. It is natural to think that they are central to copying behavior. But their presence in some primates, macaque monkeys, has not been associated with copying.

[18] The distinction between imitation and emulation is due to Michael Tomasello, 'Do Apes Ape?' in Cecilia M. Hayes and Bennett G. Galef (eds.), *Social Learning in Animals: The Roots of Culture* (New York: Academic Press, 1996), 319–346.

representations (Act)$\underline{[ego_i]that_2[ego_i]F}$ and (K)$\underline{[ego_i]that_2[ego_i]F}$ constitutes a social analog of the objectification that emerged in the mirror test.

In imitation, the individual's psychology does not achieve the full reflexivity that it does in passing the mirror test. There is no coupling of a single-indexed visual perception with a doubly indexed actional or proprioceptive perception *of the same token act*. And there is no informative coupling within vision of an ordinary perception of a seen token act with a doubly indexed visual perception *of the same token act*. Only singly and doubly indexed representations of different tokens of the same type of action are coupled.

Although imitation begins earlier in human development, it requires more, in this respect, than passing the mirror test does. Through the mirror, the couplings of seen acts with proprioceptively sensed acts are *given* to individuals that are equipped to register them. Since the individual is tracking his or her own acts visually, what is represented visually is in lock step with what is proprioceptively represented. By contrast, in imitation the individual must adjust his actions to a sequence of actions that he or she does not control. So the type matching must be made through adjustment in imitation. It is not there simply to be apprehended.

A little reflection indicates how fundamental dyadic copying, including imitation, is in infant life. Parent and child enter into games of ritualized turn taking. The range of cases in which coupling between the child's visual representations and its proprioceptive representations, for achievement of sameness of representational type, must be enormous. These couplings are the backbone of early self-objectification.[19]

In human beings, dyadic copying occurs earlier than the triadic coordination commonly involved in *joint attention*. Joint attention is widely regarded as the culmination of coordinated social interaction in the pre-linguistic phases of human development.[20] Joint attention emerges in the middle of the second year of human life. Roughly, joint attention is a capacity to share attention with another, usually toward a third object, with some mutual awareness of the shared attention.

The intermodal couplings in joint attention are complex topics in themselves. But I will not discuss their form or content here. As with passing the mirror test and imitation, joint attention does not in itself involve attribution of psychological states.[21] What is attributed to others is essentially purposive behavior. The purposiveness need not be represented as deriving from psychological states.

[19] Compare. Vasudevi Reddy, 'Before the "Third Element": Understanding Attention to Self', and Sue Leekam, 'Why Do Children with Autism Have a Joint Attention Impairment?' both in N. Eilan et al. (eds.), *Joint Attention: Communication and Other Minds* (New York: Oxford University Press, 2005), 85–109, 205–259.

[20] Malinda Carpenter, Katherine Nagell, and Michael Tomasello, 'Social Cognition, Joint Attention, and Communicative Competence from 9 to 15 Months of Age', *Monographs of the Society for Research in Child Development*, serial no. 255, 63:4 (1998).

[21] Although attending is a psychological act, and although in joint attention individuals represent acts (gazes) that are in fact guided by attention, the acts are first represented as purposive activity, not as psychologically guided purposive activity. See note 14 above.

Both imitation and joint attention do, however, provide further elements in the developmental background for robust *self*-representation.

We mature into a psychological conception of self. But we seem not to start there. Our initial ego-representations are bodily. They involve attribution of purposive activity. The activity is overt physical action. The *psychological* attributions young children come to make are, I believe, primarily built on perceptual representations of bodies, bodily agents, bodily attributes, supplemented with teleological representations of purposive or goal-directed acts. Teleology is not psychology. Our representation of ourselves as psychological beings grows from representation of ourselves and others as purposeful denizens of a physical world.

V

Intermodal perceptual coupling of ego-centric indexes is probably the most primitive representational antecedent of self-representation and self-understanding. A seemingly more advanced antecedent, which I will discuss for the remainder of this lecture, is a certain type of memory. Selves are necessarily extended in time. I think that a point-event self is an incoherent notion. Acting and experiencing take time. A self-concept must be associated with representations of oneself as extended in time. A certain type of memory underlies such representations. This type is systematically associated with a certain type of anticipation.

I will rely on a rather extensive taxonomical background.[22] Please bear with me. The delineation of types of memory is, of course, a matter for empirical psychology. I will present a classification that is a conceptually clarified version of what can be found in empirical work.

I use the term 'memory' only for a representational capacity. I intend 'representational', as usual, in my relatively demanding sense—requiring an explanatory use for appeal to veridicality conditions in type-identifying states and capacities. I do not take retentions of the effects of classical conditioning, or so-called muscle memory, to count as memory, properly so called.

Iconic memory, working memory, and other types of short-term memory have short automatic decay times and, in the latter two cases, fairly strict load capacities. I shall be mainly concerned with a certain type of long-term memory. Long-term memory encodes relatively significant matters, usually tagged through attention. Long-term memory effectively retains representational content indefinitely, with no known automatic decay times. All these types of memory occur in a wide variety of animals, including many that lack propositional attitudes.

Let us distinguish between *experiential* memory and *non-experiential* memory. Experiential memory is remembering a particular entity x in a direct

[22] This taxonomy improves on one I use at the outset of 'Memory and Persons'.

de re way. The particular can be a mental event, a physical event or physical object, or a scene. If I see rain in Salisbury and thereby remember that it rained in Salisbury, without coding the memory perceptually, my memory is not experiential, though it derives from perceptual experience. If I think *cogito* and remember that I did, but do not remember my thinking it, the memory is not experiential.[23]

The division between experiential and non-experiential memory cuts across most of the types already mentioned. Iconic memory is always experiential. But working memory, short-term memory, and long-term memory can be experiential or non-experiential. Propositional thought can occur in any of these latter three types of memory. Experiential memory can be propositional or non-propositional, as long as it is *de re* of a particular or particulars.

I will focus on a type of long-term experiential memory—*episodic memory*. *Episodic* memory is conscious, long-term, experiential memory whose perspective functions to pick out a past particular, representing it as it was at a specific past time.[24] Tulving characterized episodic memory picturesquely, as a type of mental time travel to the past.

Episodic memory evolved more recently than other memory systems. Its neural basis lies in the prefrontal cortex and other neo-cortical regions. It is more vulnerable to disease, injury, and aging than other types of memory. It seems to emerge late in child development, possibly not until 24 months, after imitation and joint attention emerge. Tulving conjectured that episodic memory is unique to humans. This conjecture is controversial. It is hard to test whether long-term animal memory of specific events is *conscious*. Something like episodic memory, not necessarily with consciousness, has been plausibly attributed to birds, dolphins, and primates.

There are two primary distinctions between generic long-term experiential memory and its subspecies, episodic memory. One is that the latter must be conscious, whereas the former need not be.[25] The other is that episodic memory must function to represent a particular *as it was at a specific past time*. Long-term

[23] To be experiential in this sense, a memory need not have the format of *perceptual* experience. It can derive from intellectual events. I am using 'experiential' in a very broad sense.

[24] The past particular could be an object or scene, as well as an event. Discovery and investigation of episodic memory as a distinctive psychological phenomenon derives from the work of Endel Tulving, *Elements of Episodic Memory* (New York: Oxford University Press, 1983); 'Remembering and Knowing the Past', *American Scientist* 77:4 (July–August 1989), 361–367; 'Episodic Memory and Autonoesis: Uniquely Human?' in Herbert S. Terrace and Janet Metcalfe (eds.), *The Missing Link in Cognition: Origins of Self-Reflective Consciousness* (New York: Oxford University Press, 2005), 3–56. Cf. also Alan D. Baddeley, Martin A. Conway, and John P. Aggleton (eds.), *Episodic Memory: New Directions in Research* (New York: Oxford University Press, 2002).

[25] Long-term experiential memory is easier to investigate empirically insofar as one does not require consciousness. There is little agreement about how to determine whether animals' memories are conscious. Often experimenters stipulate that they are studying animal '*episode-like*' memory precisely to avoid the issue of consciousness. Compare Nicola S. Clayton and Anthony Dickinson, 'Episode-like Memory during Cache Recovery by Scrub Jays', *Nature* 395 (Sept. 17, 1998), 272–274; Charles Menzel, 'Progress in the Study of Chimpanzee Recall and Episodic Memory', and Bennett L. Schwartz, 'Do Nonhuman Primates Have Episodic Memory?' both in Terrace and Metcalfe (eds.), *The Missing Link in Cognition*, 188–224 and 225–241 respectively. Tulving's characterizations of

experiential memory can be *de re* of a particular without picking out the particular episodically. A non-episodic long-term perceptual experiential memory could have a perceptual residue that is too generic, or too much a perceptual composite of experiences, to single out the particular as it was at a given time.

Both generic long-term experiential memory and episodic memory are *de re*. *De re representation* is representation that picks out a particular on the basis of a non-inferential, not purely attributional or descriptional representational competence.[26] Non-episodic experiential memory can be *de re* recognition, without representing an entity as it was at any specific time. Episodic memory is not only *de re* reference to past particulars. The *de re* representation retains a past representational state that functioned to represent a particular as it was at a corresponding past time; and the *de re* representation is accurate if and only if it derives from, and accurately represents, the particular as it in fact was, and was represented, at the relevant past time.

The notion of episodic memory, with its evocative metaphor of mental time travel, can perhaps be made vivid by contrasting it with another type of long-term memory. Imagine an animal that buries food and returns to uncover it months later. Whether the unearthing behavior indicates episodic memory—or even non-episodic experiential memory—is so far completely open, even laying aside the issue of consciousness. Consider the following updating mechanism, which is thought to occur in many non-episodic types of memory. The animal perceives the burying. A non-propositional representation with the content, that food in such and such a location now, is laid down. Memory updates this content as time passes: that [remembered] food in such and such a location now. The remembered food is expected to be currently in the location. Although memory derives from past experience, there is no memory of the burying, or of the food as it was at the time of the burying, or of the burial time itself.

Such updating use of memory does not rule out having representational content in past-tense form: that food once buried in such and such location. This non-episodic memory still lacks *de re* reference to the past. Human children younger than 24 months show no signs of episodic memory. Before that time, they can recount past events as being past. But their retellings are generic, script-like.[27] Events are situated within a script, but not on the basis of an individualized

episodic memory confine *representata* of such memory to events. But memory of bodies and scenes can be included, as long as memory represents them as they were experienced at a particular time.

[26] For discussion of *de re* attitudes, see my 'Belief *De Re*', *The Journal of Philosophy* 74:6 (1977), 338–362; reprinted with 'Postscript to "Belief *De Re*"', in *Foundations of Mind*; also 'Five Theses on *De Re* States and Attitudes', in Joseph Almog and Paolo Leonardi (eds.), *The Philosophy of David Kaplan* (New York: Oxford University Press, 2009), 246–316.

[27] Katherine Nelson and Janice Gruendel, 'Children's Scripts', in Nelson (ed.), *Event Knowledge: Structure and Function in Development* (Hillsdale, NJ: Lawrence Erlbaum, 1986), 21–46; Nelson, 'Emerging Levels of Consciousness in Early Human Development', in Terrace and Metcalfe (eds.), *The Missing Link in Cognition*, 116–141; Teresa McCormack and Christoph Hoerl, 'Memory and Temporal Perspective: The Role of Temporal Frameworks in Memory Development', *Developmental Review* 19:1 (1999), 154–182. During ages three to five, there is a leap in richness, detail, and

autobiographical experience. Human adults who have lost episodic memory through neural damage or aging can tell historical, even autobiographical narratives. But the narratives are like scripts learned from hearsay or reading. The narratives rest on memories that do not enable the individual to re-experience past events, or relive the past.

Lacking episodic memory, with its *de re* representation of particulars as they were in the past, an individual would be unable to represent such particulars in the way that the individual experienced them. The form and order of the individual's own past would be lost. For an individual with propositional attitudes, the loss would be a deficit in understanding. The individual's understanding of his or her representational past would be at best indirect.

Earlier I discussed objectification through couplings between different perceptual modalities. Such couplings coordinate representations of the same entity from the perspectives of different perceptual modalities. Such couplings provide a basis for weak ego-objectification, inasmuch as an egocentric index is attached to representation *associated* with a *perceived* entity.

This phenomenon can be replayed in perceptual experiential memory. *Ordinary*, non-proprioceptive, perceptual experiential memory—episodic or not—is analogous to visual perception in being singly indexed. It does not in itself associate an index with any remembered entity. It just carries the ego-centric index of the purported original perceptual experience, or some conceptual analog. Nothing in the remembered scene is represented as having privileged ego-status for the rememberer. There are, however, experiential memories that are the memory-cousins of doubly indexed visual perception that passes the mirror test. Such doubly indexed perceptual experiential memories are autobiographical.[28] I call an experiential memory that associates an ego-centric index, pre-conceptual or conceptual, with the individual rememberer, represented as

specificity in narration. Stories are made up as well as recounted. Temporal indexicals such as yesterday and tomorrow emerge in the third year, later than indexicals like I and you. I disagree with accounts of episodic memory, such as the last one listed, that maintain that episodic memory constitutively involves not only *de re* representation of past entities as they were at past times, but also self-consciousness and a capacity to 'interrelate arbitrary events'. Such requirements are held to be necessary to 'give meaning' to the *de re* elements in episodic memory. I believe such views commit the *philosophical* error of hyper-intellectualizing the phenomenon, along lines analogous to those that I criticize, with respect to perceptual reference, in *Origins of Objectivity*, chapter 6. The *de re* representation of past events as they were at specific past times requires a capacity to distinguish events in different instances of script-like or recurrent phase-like sequences. This capacity requires either a capacity to localize one event in a sequence, or a capacity to distinguish differences in types of events within otherwise similar sequences. But it does not require a use of allocentric, decentered frameworks, or a capacity to refer to arbitrary past events—much less self-consciousness.

[28] Episodic memory need not be conceptualized. It can be purely perceptual. Episodic memory also need not be meta-representational. It need not *attribute* psychological states, beyond the "attribution" involved in associating egocentric indexes with remembered entities, in autobiographical cases. Even when autobiographical, it may make only physical and teleological attributions to entities, marking them as having ego-significance. A common mistake among psychologists who discuss episodic memory is to claim that episodic memory constitutes a meta-perspective on one's perspective. There is no basis in the phenomenon for claiming that it is representation *of* perspective. Episodic memory

an entity in the remembered situation, an '*autobiographical experiential memory*'. An individual's autobiographical perceptual experiential memory not only functions to preserve a previous ego-centric index functioning as perceptual origin. The memory also represents the rememberer, placing the rememberer in the remembered situation; and it marks the remembered rememberer, thus represented in memory, as having ego-significance.

One way in which an autobiographical perceptual experiential memory can place the rememberer in the remembered scene is as an object of perception. An individual can remember and associate an ego-centric index with him- or herself as seen in a mirror—the memory analog of passing the mirror test. But an autobiographical visual experiential memory need not place the rememberer in the remembered scene as an object of perception. One can remember oneself *as* experiencing something other than oneself. Retaining a visual memory is distinct from retaining a visual memory in which one is also remembered as the visual perceiver. Such memory probably requires *conceptual* capacities. Perhaps most autobiographical *visual* experiential memories among selves, and *mature* persons, will be of this sort.[29]

I call a memory that is veridical, and whose veridicality includes preservation of the representational perspective of a representational state that is retained in the memory, 'memory *from the inside*'. I am interested in autobiographical episodic memories from the inside.

In the case of *perceptual* episodic memories, this notion of being from the inside is fairly straightforward. The memory retains the angle of perception—the perceptual perspective—on the subject matter that the original perception had. Not all perceptual memories are from the inside, in this sense. One could have a memory, in perceptual format, as of some subject matter. The memory could get right intrinsic properties and relations of the perceived subject matter. But the memory could be from a perceptual perspective that one never had. The memory is partly veridical, partly not. One might remember the layout of the scene quite correctly, but from a perceptual perspective outside of one's actual past perspective.[30]

retains and enables one to call up perspective. The perspective is *reused, not referred to* in the simpler forms of episodic memory.

[29] All experiential memories that place the rememberer in the remembered situation but do not do so through remembering the rememberer as an object of perception must have a conceptual, or at least post-perceptual, attributional element. Remembering oneself *as* a perceiver requires a representation as of perception. Perception itself cannot represent perception as such.

[30] A common but dramatic way in which a perceptual memory may be partly veridical, but not be from the perceptual perspective of the original experience, is for a memory to have a positional origin that is a rotation from that of the original experience. I have memories of an important childhood event from an impossible perspective twenty feet above the ground. Often a memory nonveridically places the rememberer as a perceived participant in the scene. In the psychological literature, memories from an external observer's position are called 'observer memories'. They are contrasted with 'field memories', those that take up the same positional perspective of the original experience. See Sigmund Freud, 'Screen Memories' (1899), in James Strachey (ed. and trans.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud, Volume 3* (London: Hogarth,

The notion of retaining the perspective of a memory that is non-perceptual—for example, retaining a perspective of a particular occurrent inference or occurrent thought—is similar, even if less phenomenologically vivid. Retaining the cognitive perspective of the original intellectual occurrence—and being an episodic memory from the inside—is veridically retaining the mode and representational content of that occurrence.[31] Inasmuch as the memory is episodic, it must also be a *de re* memory of that occurrence.

Thus to be from the inside, an episodic memory must be veridical, and the veridicality must include retention of the perspective of the original perceptual or intellectual "experience".[32]

I have described autobiographical episodic memory from the inside because I believe that it constitutes another step toward self-objectification. It extends the process into the temporal dimension in ways that passing the mirror test, imitation, and joint attention do not.

There is dispute in the developmental and ethological literatures over how *meta-psychological representation*—representation of psychological kinds and instances as such—emerges. I think that the matter is not well understood, and will not discuss it here. Certainly by age three or four, human children have psychological concepts and apply them to themselves and others. I believe that when married to meta-psychological representation, autobiographical episodic memory from the inside is a momentous type of self-understanding that is constitutively necessary for being a self and for being subject to norms of morality and critical reason.

When autobiographical episodic memory attributes psychological states, it contributes to specifying one's own psychological history. It enables one to revive one's psychological past in some of the same form that it took. It connects one's present and past perspectives. Unlike updating memory, it provides a basis for attributing a contrast of perspectives. When an individual can *refer* to perspectives as such, episodic memory provides a direct, *de re* way for the individual to represent him- or herself as a psychological being with changing commitments.

VI

I conclude by reflecting on how the type of episodic memory that I have been discussing gives point to the Kantian conception of self that I cited at the outset.

1953–74); Georgia Nigro and Ulric Neisser, 'Point of View in Personal Memories', *Cognitive Psychology* 15:4 (1983), 467–482; John A. Robinson and Karen L. Swanson, 'Field and Observer Modes of Remembering', *Memory* 1:3 (1993), 169–184.

[31] There may be more to the original perspective than its cognitive aspects—what Frege called 'coloring', for example. I focus on retaining cognitive aspects of the perspective.

[32] This notion of veridically retaining the original perspective must allow for more or less. Most "veridical" memory involves some omission and some errors. Exact sameness with the contents of the perceptual point of view rarely or never occurs.

Recall that Kant wrote of a being that 'is conscious of its numerical identity, of its self, in different times'. What sort of diachronic consciousness is involved? Why does it matter? I will first make some remarks about consciousness. Then I discuss the relevant consciousness's significance.

Kant is right that consciousness figures prominently in our conception of selves. The relevant consciousness is representational and meta-psychological. The representation is both of the self and of its psychological states and events.[33] It attributes psychological states, and it associates an ego-like representation of them with them.

Consciousness is a distinguishing feature of episodic memory. Meta-psychological autobiographical episodic memory from the inside retains a past perspective in an ego-like way, providing a conscious representation of one's psychological past. Insofar as such memory is consciousness of *self*, it involves consciousness of an individual's *conscious* psychological past.

The consciousness in the remembering is, I think, both phenomenal consciousness and rational-access consciousness.[34] That is, the episodic memories have a what-it-is-like quality to them; and, further, they are accessible to the individual's rational powers—propositional attitudes. Both sorts of consciousness are conditions of the whole individual, as distinguished from a psychological subsystem. Conscious psychological states that the relevant episodic memories are *of* are also both phenomenally conscious and rational-access conscious.

Perhaps not all individuals with psychologies are conscious. Some animals capable of perceptual representation may not be. But selves are conscious in both senses.[35] Phenomenal consciousness is the primitive center of conscious psychological life. Rational-access consciousness either incorporates phenomenal consciousness into propositional attitudes or consists in the individual's direct, occurrent control of propositional attitudes in mental agency.[36]

Imputation of activity or affect *to an individual*, as distinguished from attribution to the individual's psychological subsystems, motivates what has traditionally been taken as the aspect of an individual's psychology that is the individual's *own* in a proprietary sense. This notion applies, I think, to all individuals with a psychology. In higher animals, both sorts of consciousness are sufficient

[33] Kant assumes these points. In the subsequent discussion (*Critique of Pure Reason* A361–362), he takes the consciousness to be not only of self but of 'determinations' (*Bestimmungen*), which are properties of self that are of the sort that help make it a self—psychological properties.

[34] Here I use Ned Block's distinction, with a refinement that I have suggested. See Block, 'On a Confusion about a Function of Consciousness', *Behavioral and Brain Sciences* 18:2 (1995), 227–247; and my 'Two Kinds of Consciousness', in Block, Owen J. Flanagan, and Güven Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates* (Cambridge, Mass.: MIT, 1997), 427–434, also printed in *Foundations of Mind*, 383–391; 'Reflections on Two Kinds of Consciousness', in *Foundations of Mind*, 392–419.

[35] No type of episodic memory is constitutively necessary for either phenomenal consciousness or rational-access consciousness. There are certainly phenomenally conscious beings that lack episodic memory. There are probably beings, higher animals and/or young human children, that have rational-access consciousness, but lack episodic memory.

[36] 'Two Kinds of Consciousness'; 'Reflections on Two Kinds of Consciousness'.

conditions for this proprietary ownership. They are constitutive of the complex type of psychology that determines what a self is.

I believe that the sort of consciousness that Kant had in mind *and* the sort of episodic consciousness that I have been developing support a more specific point about the consciousness of self over time. The consciousness is direct and non-inferential—hence *de re*. Kant would not have counted conscious theory about one's past, based on *inference*, sufficient for *consciousness of self* in the relevant sense. Nor would he have counted memory that is a mere summation of past experiences *consciousness of self* in the relevant sense.[37]

Meta-psychological, autobiographical *episodic* memory *from the inside* meets this requirement of a *de re* relation to the past—and illuminates it. In meta-psychological uses, episodic memory represents *instances* or *tokens* of past psychological events or states. This sort of memory contrasts with two other kinds of meta-psychological autobiographical memory. One type is generalized memory of types or patterns that is an amalgam or summation of individual past experiences, but that does not retain any one past psychological event or state instance and represent it as it was at a specific past time. Memories with the following contents are examples: my experiences with La Tâche have been thrilling; I have often seen him dominate her; I have learned that she is not to be trusted. The other type is memory that characterizes a specific event, but in a generalized way that does not involve *de re* memory *of* the event. Memories like the following are examples: I have once seen a nut buried there; I have felt this sort of pain once before; at some point, I came to believe that Schubert's string quintet is sublime.

None of these memories is in itself a memory *of* a particular event instance. The memories either summarize several past events without necessarily remembering one of them as it occurred, or they retain *that* a particular event occurred without necessarily remembering that event as it was at a specific past time. Episodic memories are not only non-inferential. They are directly *of* the remembered entity. They are *de re*. I think that meta-psychological, autobiographical, episodic memories from the individual's past perspective provide the sort of *self-consciousness over time* that, on the Kantian conception, is constitutive to being a self.

I turn now to why the diachronic aspect of this self-consciousness matters. What values attach to a self conceived in this way? I think that the conception figures centrally in the applicability and application of norms that are constitutive to the natures of selves and persons.

Locke took persons to be approximately what I call selves.[38] He characterized persons as thinking, reflective, self-conscious beings capable of extending

[37] Kant takes the consciousness to be via inner *sense*, which is non-inferential: *Critique of Pure Reason* A361–363. I agree with Kant that the consciousness is non-inferential. I do not agree that it must be via a *sense*—a capacity that yields only empirical cognition. See note 1.

[38] Locke took persons to be a subspecies of what *he* counted selves. He explained selves thus: '*Self* is that conscious thinking thing, (whatever substance, made up of whether spiritual, or material, simple or compounded, it matters not) which is sensible, or conscious of pleasure and pain, capable of happiness or

consciousness to different times and places.[39] Locke regarded person as a 'forensic' notion, one essentially associated with responsibility.[40] Locke's first characterization of persons probably motivated Kant's focus on diachronic self-consciousness. Locke highlights extending consciousness 'backwards' in time to 'action or thought'. (See notes 39–42.) This emphasis is motivated by a concern that persons be able to understand their responsibility for past misdeeds for which they could be punished.

Locke gives no account of extension of consciousness into the past, although he clearly takes instances of such extension to be memories. His discussion of accountability, combined with his writing of *repeating* past consciousness in extending consciousness into the past,[41] suggests that he may have had episodic memory in mind. I shall assume that he did.[42]

I believe that Locke was right to think that the relation between moral accountability and episodic memory is closely connected to the psychological *kind* that selves instantiate. However, he took the connection to hold at too specific a level. I will discuss these matters in more detail in the next lecture, but I sketch some issues here.

An individual can be morally responsible for deeds that he or she does not remember. An individual can be responsible for deeds that are motivated from a perspective that he or she does not know introspectively at the time the

misery, and so is concern'd for it *self*, as far as that consciousness extends.' *An Essay Concerning Human Understanding* (Oxford: Clarendon Press, 1975), Bk. 2:27: 17. See also 2:27: 25–26. I call 'egos' *roughly* what Locke calls the selves of beings that are not persons. I do not require an ego to be conscious. Egos must be purposive and have some representational competencies, at least perceptual ones. Locke's persons are selves (in his sense of 'self ') that are intelligent, have rational reflection, can extend their consciousness to different times and places, and are morally accountable. Locke's persons are *roughly* what I count selves. I say 'roughly', here and above, partly because I do not agree with Locke on the identity conditions for persons, selves, or egos. Some of these differences will emerge shortly.

[39] Locke's fuller characterization of person: '. . . a thinking intelligent Being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places, which it does only by that consciousness which is inseparable from thinking, and as it seems to me essential to it . . . . in this alone consists *personal identity, i.e.* the sameness of a rational being. And as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that *person*; it is the same *self* now it was then; and 'tis by the same *self* with this present one that now reflects on it, that that action was done.' *Essay*, 2:27: 9.

[40] For the claim that 'person' is a forensic term, see *Essay*, 2:27: 26. Locke develops his views on persons and moral responsibility in 2:27: 18–26. For an illuminating discussion, see Patricia Kitcher, 'Two Normative Rules for Self-Consciousness', in Terrace and Metcalfe, (eds.), *The Missing Link in Cognition*, 174–187.

[41] In *An Essay Concerning Human Understanding*, 2:27: 10, Locke writes of persons' *repeating* 'the idea of any past action with the same consciousness it had of it at first'. This passage makes a relatively close fit with meta-psychological autobiographical episodic memory.

[42] It has been disputed whether Locke understands 'extension' of consciousness back in time to be a type of memory. Some have maintained that he meant merely continuity of consciousness. I think that his taking this extension to be something a person does (*Essay*, 2:27: 9–10) and his viewing personal identity and moral accountability as hinging on memory (*Essay*, 2:27: 20–22) suggest that he views extension of consciousness into the past as memory. At least, he takes memory to be a primary *type* of such extension.

motivation is formed. Contrary to Locke, a drunkard can be morally responsible for misdeeds that are neither remembered nor understood at the time.[43] People mired in self-deception or naiveté about their own psychologies are similar cases in point. Moreover, longevity of life can lead to forgetting misdeeds for which one remains responsible. It is simply not true that one must have an autobiographical episodic memory, or indeed any other memory, for each past deed for which one remains morally responsible.

The connection of accountability with meta-psychological, autobiographical episodic memory holds at a more general level. As a first approximation, I think that having a natural competence consciously to revisit some actions and state instances from the inside is a competence that is constitutive to our *status* as morally accountable beings.

Darwin points in the same direction. But like Locke, he underestimates the subtlety of the connection between memory and being subject to moral norms. He writes:

> A moral being is one who is capable of reflecting on his past actions and their motives—of approving of some and disapproving of others; and the fact that man is the one being who certainly deserves this designation, is the greatest of all distinctions between him and the lower animals.[44]

Although Darwin shares Locke's focus on memory, he aims—fruitfully, I think—at a generic connection between memory and moral status, rather than a connection between memory and each morally responsible act. On the other hand, Darwin is less focused on episodic memory than Locke is. Furthermore, approval and disapproval are not enough to make an individual a moral being, even when added to diachronic competencies that are relevant to moral status. An individual must also understand a distinction between right and wrong. Nevertheless, I think Darwin right to connect diachronic competencies with being subject to moral norms.

The constitutive connection between being subject to moral norms and being competent to extend consciousness over time is not that meta-psychological autobiographical episodic memory is the link that enables a person or self to remain the same person or self over time, as Locke seems to have thought. Identity is a necessary condition on primary moral accountability. But episodic memory is not the primary link in transtemporal identity. There are many other

[43] Locke's claims, now widely rejected, occur in *Essay*, 2:27: 22–26. Leibniz, *New Essays on Human Understanding*, ed. and trans. Peter Remnant and Jonathan Francis Bennett (New York: Cambridge University Press, 1981), 2:27: 9, criticizes Locke's views on accountability and personal identity. Leibniz allows for memory gaps in both. He maintains, correctly, that testimony of others can suffice to enable one to understand the justice of punishment for unremembered deeds. Of course, it does not follow that if all that one ever had had—even for very short-term time spans—was testimony from others about one's past, one would be a morally responsible person.

[44] Charles Darwin, *Descent of Man* (1871) (Amherst, NY: Prometheus, 1998), 633.

psychological links, including unconscious ones and ones other than memory, that help make possible the identity of selves and persons over time.

The connection between diachronic competence and moral accountability is more abstract. In lacking a competence for transtemporal 'extensions of self-consciousness'—of which a competence for meta-psychological autobiographical episodic memory is a central example—an individual lacks an essential aspect of a certain *understanding*. A capacity to understand one's actions in this way is necessary for being subject to moral norms, and partly constitutive of being a self. My next task is to explore the kind of understanding that has these features—with special attention to the role that memory plays in this self-understanding.

8 *Self and Self-Understanding*

Lecture II: Self and Constitutive Norms

In this lecture, I pursue a certain methodological strategy. I want to contribute to understanding the nature of selves. I think that certain sorts of self-understanding are among the constitutive features of selves. I explore that type of self-understanding by using clues from requirements on the applicability of two types of norms—those of critical reason and morality. Being subject to norms of critical reason is plausibly constitutive to being a self. Being subject to moral norms is at least arguably constitutive. Having certain psychological competencies, including relevant sorts of self-understanding, is constitutive to being subject to these norms. So the requirements for being subject to these norms provide insight into the nature of the relevant self-understanding, and ultimately into the nature of selves and persons.

I

A *norm* is a standard that governs good or optimal ways of fulfilling some function, or achieving some goal, end, or value.[45] Many norms do not require for their applicability any sort of understanding on the part of individuals that are subject to them. Norms for having a given standard of living or for having a life-sustaining nutrition, for any organism, are applicable even if they are not understood. Norms for the well-functioning of the heart or for formation of veridical visual perceptions require no understanding. They apply whether or not anyone knows about them. Norms are applicable given any function. Norm is a teleological notion. Not all teleology involves psychology, much less understanding.

Even some *rational* norms governing propositional attitudes require neither self-understanding nor understanding the norms. Reasoning by young children is subject to norms of deductive and inductive inference. Being subject to these norms requires no understanding beyond competence to make such inferences.

[45] See my 'Perceptual Entitlement', *Philosophy and Phenomenological Research*, 67:3 (November 2003), 503–548; and 'Primitive Agency and Natural Norms', *Philosophy and Phenomenological Research* 74:2 (September 2009), 251–278.

An individual is subject to these norms if it engages in inference. Inference has the representational function of preserving truth, supporting truth, or supporting some goal. Norms of reasoning apply to any being with propositional attitudes. Organisms that lack propositional attitudes can be evaluated for efficiency, but not reasonability. But young children and probably some non-human animals are subject to norms of reason, even though they cannot understand the norms.

I shall center on moral norms and norms of critical reason. Moral norms are norms for acting morally well, for making good decisions, for reasoning well to moral conclusions, and for having good character. Norms of critical reason perhaps need more explication.

Critical reason is a capacity to recognize and effectively employ reasonable criticism of or support for propositional attitudes and for propositional reasoning, guided by an appreciation, use, and assessment of reasons and reasoning as such. Critical reasoning is the sort of reasoning that evaluates, checks, refines ordinary first-level reasoning, using concepts and standards of reason. Critical reason is constitutively a capacity for meta-reasoning and meta-evaluation. It is commonly reasoning about ordinary subject matters and about reasons at the same time. It evaluates attitudes and reasoning for their reasonability.

Most of our reasoning, even most of our most sophisticated reasoning, proceeds at a first-order level. Most of our reasoning does not refer to representational contents or psychological states, and does not invoke the concept reason or standards for being reasonable. It simply operates under those standards more or less well. Critical reasoning is not in itself any better or more effective as reasoning than first-order reasoning. *Actual use* of critical reasoning is not necessary for making scientific discoveries. But a *capacity* to frame and defend a methodology is constitutive of science. Entering fully into practices that allow us to have our culture and history requires some capacity to explain and defend the practices. One must be able to answer why questions if one is to stand behind—with good or spurious reasons—the values, taboos, and permissions, the institutions and relationships, that make up a civilization. Such ability constitutes a rudimentary capacity for critical reasoning.

We are probably not the only species that reasons. Other animals probably carry out propositional, means–end inferences, subject to norms of reason. Ability to reason is probably not distinctive of selves or persons. But I think that ability to reason *critically* is.

I will not rely on a view about the relation between norms of critical reason and those of morality. Some, like Hume, hold that moral values are a matter of non-rational choice—a rationally optional project.[46] Others, like Kant, hold that moral values are matters of reason. Some think that a person or self could be irretrievably and naturally morally blind or indifferent—without damage, deprivation, illness, or malfunction. Others, following Aristotle and Kant, maintain that

[46] Even Humeans can allow a constitutive role for critical reason in morality—a role for a capacity to defend choices, including moral ones, as reasonable, *relative* to one's moral values.

moral values are associated with constitutive functions of human beings, persons, or selves. I am more sympathetic to the latter in each pair of views. Since the self-understanding that I discuss is parallel for morality and critical reason, little in the discussion will depend on either of the issues cited in this paragraph.[47]

I think it hardly controversial that having a natural capacity for critical reason is constitutive to being a person and being, or having, a self. It is arguable that being subject to moral norms is also constitutive. I will focus on both kinds of norms as clues to the nature of a type of self-understanding that is constitutive to being a person and self.

An individual's instantiating certain psychological kinds—including a kind of self-understanding and an understanding of notions like reason and moral wrong—is constitutively necessary, and arguably sufficient, for the applicability of the norms of critical reason and morality to that individual. Reciprocally, these psychological kinds necessitate the applicability of certain norms to their instantiations. So psychological capacities and normative applicability are interdependent. I shall use the applicability of these norms as clues to the natures of underlying psychological capacities.

The norms of morality and critical reason are not typical norms. Unlike norms for good nutrition, cardiovascular health, and veridical visual-perception formation, these norms apply only to beings with certain psychological capacities for propositional attitudes. The norms' applicability hinges on two other requirements. They require *understanding the norms*—some minimum internalization of them. And they require some *self-understanding*—an understanding of psychological conditions, in one's own case, to which the norms apply.[48]

[47] Locke's claim that person is a forensic notion expresses his view that there is a constitutive connection between being a person and being responsible to moral norms. This view is independent of Kant's more committal view that morality rests on critical reason. Hume differs from Kant in holding that morality does not depend on reason's determining basic values. I do not know whether he thought that a capability for morality is necessary to being a nondamaged fully mature person, or human being.

It is not uncommon to take Locke's claim to imply that the kind *person* does not mark off individuals and is not a basic psychological kind. Locke understood a forensic notion as one having to do with judicial proceedings and law, including divine moral judgments. It is, even now, common to contrast forensic notions with "ontological" notions like substance. Such a contrast is often taken to imply that the kind *person* marks a stance, phase, or contingent property—not a kind of individual. I believe that such contrasts often depend on an outmoded notion of substance. The stance-oriented way of thinking about normative notions also tends to conceive norms as applicable only given desires or choices. There are some norms of this sort. Perhaps norms of table setting apply only if one cares to conform. But norms of critical reason are not like that. And neither, I think, are norms of morality. I believe that these norms apply to individuals if and only if the individual has certain psychological capacities and certain rudimentary understanding. At any rate, I think that counting person and self notions with normative implications does not *entail* that they apply to less ontologically significant kinds.

[48] Tyler Burge, 'Reason and the First Person', in Barry C. Smith, Crispin Wright, and Cynthia Macdonald (eds.), *Knowing Our Own Minds: Essays on Self-Knowledge* (New York: Oxford University Press, 1998), 243–270.

Let us briefly consider the requirement of understanding these norms.[49] Being subject to moral norms requires some understanding of the difference between right and wrong. Both right and wrong are moral, normative notions. Right and wrong centrally concern psychological motivations and psychological initiators of action. The motivations and initiators must involve conceptualization of actions. Application of the norms to actions requires that actions be individuated partly by their psychological antecedents. Helping an old lady across the road mainly motivated by a desire that she arrive at the bank where a scam is planned is a morally different act from helping the old lady across the road mainly out concern for her well-being.[50] Minimal understanding of how moral

[49] Development in children of a capacity to apply norms is not deeply understood. Here is a rough summary. Child development from ages three to six is marked by changes in understanding the role of norms in self-regulation. By age three or four, when children clearly exhibit an ability to attribute beliefs and intentions, they show some awareness that others have different beliefs and goals. Between ages three and six, children exhibit an understanding that social regulation by others is mediated by others' attitudes. Evaluations had been applied to appearance, physical prowess, social behavior. They come to be applied to psychological states. Evaluation is initially accepted from others. It comes to be self-applied. Between ages six and ten, children come to evaluate individuals' psychological conditions and characters not just in terms of whether an outcome matches an antecedent intention or belief. They evaluate individuals in terms of the quality of the antecedent state, somewhat independently of whether an outcome is favorable. This distinction probably grows out of a distinction between talent and effort in physical pursuits. Such evaluations have a three-factor profile. First, there is the contribution of the individual's abilities (those relevant to talent, character, psychological capacity) to whether a standard is met. Second, there is the contribution of the individual's use of these abilities. Third, there is the contribution of external conditions to success or failure. Moral evaluations of individuals and their acts center on psychological capacities, character, and use of them. An individual can act well morally, although things turn out badly. An individual can act badly, although things turn out well. Similarly, evaluations for critical reason center on the individual's contributions, not on extra-psychological matters. Such evaluations are usually independent of the veridicality of an attitude, and of the success of the activity in realizing the individual's goals. An individual can be rational but mistaken, or right but irrational.

For a sampling of literature on development of cognition of norms, see Henrike Moll and Michael Tomasello, 'Level 1 Perspective-Taking at 24 Months of Age', *British Journal of Developmental Psychology* 24:3 (September 2006), 603–613; Charles W. Kalish and Sean M. Shevcrick, 'Children's Reasoning about Norms and Traits as Motives for Behavior', *Cognitive Development* 19:3 (July–September 2004), 401–416; Nancy Eisenberg, Richard A. Fabes, and Tracy L. Spinrad, 'Prosocial Development', in Eisenberg, William Damon, and Richard M. Lerner (eds.), *Handbook of Child Psychology, Volume 3: Social, Emotional, and Personality Development* (New York: Wiley, 1998); Adam Rutland et al., 'Social Norms and Self-Presentation: Children's Implicit and Explicit Intergroup Attitudes', *Child Development* 76:2 (March/April 2005), 451–466; Gergely Cisbra and György Gergely, 'Social Learning and Social Cognition: The Case for Pedagogy', in Yuko Munakata and Mark H. Johnson, (eds.), *Processes of Change in Brain and Cognitive Development: Attention and Performance XXI* (New York: Oxford University Press, 2006), 249–274; David R. Shaffer and Katherine Kipp, *Developmental Psychology: Childhood and Adolescence* (Belmont, CA: Wadsworth, 2009).

[50] I believe that I disagree here with Derek Parfit and Tim Scanlon, both of whom minimize motive in determining the moral goodness of an action. See Derek Parfit, *On What Matters*, vol. 1, ed. Samuel Scheffler (New York: Oxford University Press, 2011); and Thomas M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge Mass.: Harvard University Press, 2008). For a view that I find more congenial, see Barbara Herman, 'A Mismatch of Methods', in Parfit, *On What Matters*, vol. 2, ed. Samuel Scheffler (New York: Oxford University Press, 2011), 83–115.

notions apply requires not only *having* such notions, but being able to use them to evaluate psychological conditions.

Similarly, being subject to norms of critical reasoning requires minimal understanding of norms of critical reasoning—understanding what a reason is. The minimal understanding involved in having the concept reason requires understanding applications of the concept to *instances* of reasoning.[51] Reasons apply centrally to representational contents taken with their mode (belief, suspension of belief, intention). But, as noted, understanding the concept reason requires a capacity to apply it to instances. Understanding conditions under which the concept applies in actual reasoning requires a capacity to apply it to psychological states—instances as well as types—as individuated by their modes and representational contents.

Thus to be subject to either moral norms or norms of critical reason, an individual must have relevant normative concepts and abilities to apply such concepts competently to psychological states and occurrences, and their representational contents, as such. These are *higher-order*, ability-general metapsychological capacities.[52]

My main interest here is the requirement, for the applicability to an individual of norms of morality and critical reason, that the individual have self-understanding. To be subject to norms of morality, one must understand application of concepts of right and wrong to psychological states or to actions individuated in terms of psychological antecedents. This understanding must be applicable to one's own case. One must be able to evaluate one's own psychological states, particularly inasmuch as they determine and help individuate one's bodily actions; and one must be able to act on such evaluation. These abilities require an ability to understand what one's psychological states are.

Similarly, norms of critical reason are applicable to an individual only if the individual has a standing competence, at least in some cases, to understand his or her point of view and to criticize and affirm that point of view by reference to reasons considered *as* reasons. In exploring these requirements, I shall be guided by the idea that there must be a capacity to understand one's point of view and the applicability of normative concepts to it "*from the inside*". Like memory from the inside, understanding from the inside must get the understood state's representational features right, and must do so by preserving those features in certain ways. Preservation is a fundamental connector in any point of view. Selves capitalize on this connector. Conditions for applicability of norms of critical reason and morality center on understanding that makes essential use of preservation.

These points about baseline conditions for the *applicability* of the norms are fairly straightforward, as far as they go. I think that we can gain insight into the specific nature of the required self-understanding by reflecting on conditions for

[51] Indeed, instances of reasoning in one's own case. See my 'Reason and the First Person'.

[52] For explanation of the ability-general/ability-particular distinction, see my 'Five Theses on *De Re* States and Attitudes'.

the *application* of the norms in particular instances. The patterns of evaluations in such applications can be very complex, especially in the moral case. But some simple points about how antecedent psychological states and occurrences figure in such applications will, I think, lead us a step deeper.

II

Let us first look at some applications of moral norms. Moral norms apply most conspicuously to an individual's acts individuated in terms of psychological antecedents. Individual responsibility for such acts is centered on psychological antecedents. An individual can be just as responsible for a bodily act as for psychological antecedents that initiate it. But in such cases full, unmitigated responsibility for the act derives from the act's proceeding from its psychological antecedents in a natural way. The root of responsibility for the act is the individual's responsibility for psychological antecedents—the cognitions, intentions, decisions, and psychological initiations—that are subject to a certain sort of self-understanding. Moral responsibility has its constitutive, explanatory basis in psychological antecedents of acts, antecedents that are subject to a certain sort of self-understanding.

Similarly, an individual is morally responsible for consequences of his or her acts only if the consequences are traceable to something that the individual intentionally did or omitted, where the individual was capable of understanding the intention and initiation of his act, or omission, from the inside, and recognizing the act's moral implications.

On the other hand, an act can be wrong and an individual can be morally responsible for it, even if the individual does not specifically intend, or could not foresee, its harmful consequences. An individual can be morally responsible for acts carried out at times when the individual cannot understand what he or she is doing, in morally relevant terms. I noted the classical case of the drunk last lecture. An individual wild with anger can be morally responsible for injuring another, even if the injury was not specifically intended and even if the perpetrator was not in a state to understand the recklessness of the behavior. The individual might not even realize that he or she was angry or had reckless intentions. The individual could still be morally responsible for the behavior and the injury—not just for not knowing what he was doing and not just for not foreseeing the injury. Similarly, arrogant behavior is often not backed by specific intention or self-understanding. Racist beliefs are often not understood to be racist. Insensitive ingratitude or sexist acts are often not intended or understood as such. Yet an individual can be morally responsible for these states or acts.[53]

[53] Robert Merrihew Adams, 'Involuntary Sins', *The Philosophical Review*, 94:1 (January 1985), 3–31. Several of my examples are drawn from Adams's provocative article. He argues against always tracing responsibility in these cases to earlier culpable willings and understandings. I am not persuaded.

Criticism of the acts and the individual is often *mitigated* in such cases. The drunk driver is not morally as bad as the driver who lies in wait. An inappropriately angry person who would be horrified by his act in a cooler moment is often less bad than one who calculates injuring. Individuals are not strictly morally liable for all causal consequences of their acts. But neither are they always absolved of moral responsibility because they lack the self-understanding to evaluate them as they are formed.

I think that self-understood psychological antecedents in carrying out acts do, however, have a special place in *application*—not just *applicability*—of moral norms. The pattern of mitigation just cited suggests this point.

Responsibility by the drunk person hinges on the relation of the act to a baseline condition in which the individual controls his or her intentional acts and is in a position to understand their psychological antecedents and the acts' moral significance. At some time before the driver became drunk, perhaps in the development of drunken habits, the individual must have had a capacity to foresee risks to others, if the individual is to be counted morally responsible for the act. At some such time, the drunkard was in a position to understand and evaluate psychological antecedents to acts that in some foreseeable way led to being drunk.

Similar points apply to the cases of anger, arrogance, insensitive ingratitude, racism, and sexism. If these characteristics were to have developed in such a way that at *no* point was there *any* capacity to understand, evaluate, and control relevant psychological antecedents, the individual would not be morally responsible for them. It would be as if the states were injected into the individual. We control only little of our psychologies at any given time. But over time, we have opportunities to shape them. Such opportunities make possible moral responsibility for more than we control or understand at any given time.

The mitigation in moral evaluation of the drunk driver in comparison with the calculating murderer derives from a distance between the driver's condition at the time of the accident and the driver's condition when he was in a position to understand what he was doing. The fact that the driver had less understanding of and control over his point of view and its moral consequences at the time of the accident than at an earlier time is *commonly*—I do not say 'always'—a source of mitigation. Moral evaluation of individuals ultimately centers on ways an individual's acts and states relate to uses of the individual's point of view that *are* under intentional control and capable of self-understanding. Moral evaluation takes such uses as a baseline for determining mitigation.

The *application*, as distinguished from *applicability*, of norms of *critical reason*—especially theoretical critical reason—is less complex than application in the moral case. But again, ultimately, certain elements in an individual's point of view are privileged in evaluations. Let me lay out some of these differences and similarities step by step.

Norms of critical reason include norms of first-order rationality—those that apply to any reasoner—together with those rational norms that are specific to critical reasoners. Norms specific to critical reason apply only to propositional states and events that are in principle accessible to an individual's rational powers, immediate self-understanding, and rational evaluation, using the concept reason or some variant. Other aspects of an individual's psychology can be unreasonable, but not specifically *critically* unreasonable. However, elements of a point of view that the individual does not control, access, or understand fall under norms of critical reason. Elements in the Freudian unconscious, elements mired in self-deception or inebriation, elements distorted by entrenched irrationality, and elements camouflaged from self-understanding by anger or depression are subject to norms of critical reason. In this respect, application of moral norms and norms of critical reason are similar.

A difference is that norms of critical reason apply with equal force to all elements of an individual's point of view to which the norms are applicable. In the moral case, intentions and actions not presently understood or available to self-understanding are often *not* just as morally bad as they would be if they had been understood. The norms of critical reason include norms of ordinary non-meta-representational reason. Unreasonable elements in a point of view that are not presently available to self-understanding and self-evaluation are just as unreasonable as they would be if they were. The drunkard's inferences, which are not immediately accessible, but fall under standards of reason, are just as unreasonable as a sober person's making the same inferences under reflectively ideal circumstances. Some psychological states or transitions that would normally be open to self-understanding might be influenced by disease. Their unreasonability need not be explained in terms of something the *individual* did or did not do in allowing them to lapse, or in failing to bring them under the control of critical reason. Unlike immorality, unreason—even specifically critical unreason—can be *purely* the product of disease or malfunction. In assessing *failures* of reason, there seems to be no privileging of elements that are presently open to self-understanding over those that could be open to self-understanding, but are not.[54]

Still, as in the moral case, reasoning that is easily accessible to current self-understanding is privileged in *successes* of critical reason. Successes are constitutively open, readily open, to critical reflection that includes self-understanding. Psychological elements that are not currently easily accessible to self-understanding can fulfill norms of critical reason only via inertia: they were readily open, and have maintained their reasonability.

[54] There is, however, a pattern of mitigation that resembles the moral case. Suppose that an individual's reasoning is periodically clouded by anger. Suppose that the individual tries to change matters for the better, but does not succeed. The unreasonable states are just as unreasonable as if the individual had done nothing to combat them. But the *individual's* critical unreasonability can be mitigated. The *individual* is perhaps less critically unreasonable than if he or she gave in to the unreason.

An individual can use critical reasoning to operate on elements of the point of view that are resistant to critical reason, and—although they may be in principle accessible—may not be currently, easily accessible to self-understanding. We say that the individual *tries to get him- or herself* to think or act more reasonably. The recalcitrant elements are not, however, part of the critical reasoning. They are within the individual's point of view, and are subject matters of higher-order reasoning. But the aspects of the point of view that are the topics of the reasoning are not being used in critical reasoning. They are like objects to be manipulated. The resources of successful critical reasoning constitutively involve elements that are currently, easily accessible to self-understanding.

Meeting the norms of critical reason and morality occurs paradigmatically—I think, only—in cases identical with or derivative from cases in which an individual is currently in a position to readily understand the relevant norms and the application of the norms to relevant psychological states and occurrences. Understanding such application requires that the psychological elements be currently, readily accessible to self-understanding.

Let me summarize my points about applicability and application of the two types of norms. Both types center on psychological elements that are imputable to the individual subject to the norms—to that proprietary part of the individual's psychology in the individual's point of view. What happens in the environment, in intentional bodily actions, or in aspects of the individual's psychology that are irretrievably beyond the individual's control or consciousness, or that do not accord with the individual's self-understanding, have a fundamentally different status in the norms' applicability and application than elements within the individual's point of view that are subject to self-understanding.

III

Animals that are not persons and lack selves have points of view. Even in animals' points of view, there is a distinction between elements that are occurrently conscious, or are immediately or easily accessible to consciousness—whether phenomenal consciousness or rational-access consciousness—and elements that are not. Thoughts or perceptions that are retained in memory, but that would need extensive priming to gain consciousness or be available for rational use, fall in the latter category. Standing beliefs or memories that are available for use and control as soon as the occasion requires fall in the former category.

I call the part of an individual's point of view that is rational-access conscious, or immediately or easily accessible to such consciousness, '*the individual's rationally accessible point of view*'. Elements not immediately or easily accessible but imputable to the individual are part of the individual's point of view, but not part of the individual's rationally accessible point of view. Perceivers that lack propositional attitudes have a point of view, but no rationally accessible

point of view.[55] Probably all individuals with rationally accessible points of view have further elements in their points of view, outside the rationally accessible point of view. They probably have memories that need priming to come into play.

For persons and selves, this distinction between point of view and rationally accessible point of view has a higher-level analog. There is a distinction between an individual's representational states and events that are immediately or easily accessible to conscious *self-understanding*, on one hand, and representational states and events that are imputable to the individual, part of the individual's point of view—and perhaps even in principle accessible to self-understanding—but not immediately or easily accessible, on the other. I call the former representational states and occurrences, together with conscious sensations and feelings that are also immediately or easily accessible to self-attribution, the individual's '*apperceptive rationally accessible point of view*'.[56] An individual's unrecognized anger, Freudian unconscious, self-deception, and other states that need extensive priming to be recognized, are outside the individual's apperceptive rationally accessible point of view.[57] The apperceptive rationally accessible point of view is the part of a point of view that is immediately or easily accessible to conscious self-understanding.

Our reflection on applicability and application of moral norms and norms of critical reasoning indicated that apperceptive rationally accessible points of view have a privileged status. Applicability and fulfillment of these norms require a capacity for self-understanding of relevant psychological elements and a capacity to evaluate such elements under the norms. Apperceptive rationally accessible

[55] In individuals with a rationally accessible point of view, there is a further distinction among elements of the point of view that are outside that part of their point of view. There are "outside" states imputable to an individual that are not accessible to consciousness even with priming, as well as those that are. Imputability to the individual is not fundamentally a matter of accessibility-in-principle to either sort of consciousness. Individuals have short-term, unconscious perceptual beliefs and memories that enter into individual-level explanations of an individual's activity, but that are forever beyond conscious control or phenomenal awareness. I think that such beliefs are imputable to the individual, first, because all beliefs are imputable and, second, because they can enter into individual-level explanations of action. This point will have an analog in discussion of critical points of view: some psychological states that are imputable to an individual cannot be claimed by the individual, even in principle, as the individual's own and are not under control of the individual's agency. (Of course, such states tend not to be *morally* imputable.) In other words, psychological imputability and point of view have their roots in committal states and activity not in consciousness. Committal perceptual states and activity are present in the animal world at very low phylogenetic levels. These points are partly terminological. But terminological choice has important substantive implications for how we think about psychology. Whatever the terminology, these distinctions should be drawn.

[56] In persons and selves, anything in the apperceptive rationally accessible point of view is in the rationally accessible point of view. The converse may also be true: anything in the rationally accessible point of view is, in persons and selves, in the apperceptive rationally accessible point of view. I leave this issue open to further investigation.

[57] *Occurrently* conscious elements in apperceptive rationally accessible points of view have a priority in our *sense* of life. I leave open whether they have normative priority. For purposes of the large normative considerations that we are discussing, it is the full rationally accessible point of view that is privileged.

points of view set standards for successful realization of the norms. They form a baseline for mitigation in negative moral evaluation.

I term this fundamental status of apperceptive rationally accessible points of view in explanations of normative evaluation in morality and critical reason their '*buck-stopping status*'. There are no further or deeper grounds for evaluating individuals' contribution to their accountability under norms of morality or critical reasoning. Applications of these norms are rooted in individuals' apperceptive rationally accessible points of view.[58]

The idea behind this point is that elements in a point of view that have buck-stopping status with respect to norms of morality and critical reason are constitutive starting points for expression of an individual's moral being or critically reasonable self. They mark ultimate unmitigated responsibility to the norms. All other elements in an individual's point of view are at best mediate expressions. They get their normative status derivatively—through relations to the apperceptive rationally accessible part of the point of view.

The primary relations here are inferential ones. The buck-stopping status of apperceptive rationally accessible points of view lies largely in their being baseline starting points for inferences. These baseline starting points necessarily are immediately, non-inductively accessible, at least within the conscious apperceptive point of view. 'Easiness' of access is meant to be unspecific. Easiness is relative to the starting points that are immediately accessible to self-understanding. In having easy access to some elements in a point of view by inference, one must self-understand the starting points of the inference.

Although both moral evaluation and evaluation of critical *practical* reason usually center on bodily intentional action, the baseline, buck-stopping starting points for such evaluation are the psychological antecedents of the actions. Let me make this point somewhat more precise. I shall henceforth call intentional bodily action just '*action*'. As noted, one can be equally accountable under the two types of norms for action and for its psychological antecedents. The capacity to reflect on one's actions, especially prospectively, is a factor in relevant normative evaluation. My claim is that the applicability and application of the norms of morality and critical reason to actions goes through their applicability and application to the psychological antecedents of the actions. The individual's contribution to his or her accountability for actions derives from the actions' relations to self-understandable cognitive and motivating psychological antecedents. Full accountability for actions depends on ability to self-understand psychological antecedents.[59] Where understanding actions themselves is a source of

[58] I think that not all self-understanding that grounds accountability under the relevant norms is agential. I discuss relations between rational-access consciousness—the lower-level underpinning of an apperceptive rationally accessible point of view—and individual agency in 'Two Kinds of Consciousness'.

[59] There is some question about whether individuals always know directly the initiation of a bodily action—when and whether the final choice or action initiation is made. Compare Masao Matsuhashi and Mark Hallett, 'The Timing of the Conscious Intention to Move', *European Journal of Neuroscience*

the individual's contribution to his or her accountability, the understanding of the actions must be through understanding their cognitive and motivating psychological antecedents. Basic, full, unmitigated accountability for an action depends on the action's realizing the content of those antecedents, where the antecedents can be self-understood. If the action is not individuated by the self-understandable psychological antecedent—for example, if it does not realize an antecedent intention—the evaluation of the act is mitigated, or at least depends partly on its relation to actions that *are* successful expressions of self-understandable psychological antecedents. The starting point and ultimate source for evaluation of individuals under the two norms is the set of cognitive or motivating psychological antecedents. These help individuate the actions that are evaluated. They obtain their buck-stopping status partly through their potential for being self-understood.

The converse points do not hold. The basis of accountability does not lie in understanding the cognitions and motives through understanding the individual's actions. We often do understand our motives retrospectively, from the actions inward. But such understanding does not ground *fundamental* accountability of individuals under the relevant norms. Motives that are understandable only outside-in are less fundamental expressions of an individual's moral and critically rational being. And actions that derive from such motives are often evaluated as involving mitigated accountability. Basic, unmitigated accountability traces to cognitions and motives that are subject to potential self-understanding in advance of, or during, actions that they produce. The *fundamental* basis for evaluating individuals' contribution to accountability under the two primary norms is the self-understandable cognitive and motivating psychological antecedents, not the self-understandable actions.[60]

28:11 (November 2008), 2344–2351. As long as the individual is capable of understanding and knowing, in a relevantly direct way, the motivating intention, the conditions on self-understanding of an action *in* doing it are met. If the action initiation flows naturally from an intention and decision and is not a surprise, the individual can be fully accountable for his or her action, under the relevant norms, if the intention and decision are open to self-understanding, even if the final occurrent action initiation is not. The latter might well be unconscious.

[60] These points must be understood in the complex context of moral evaluation discussed earlier. Individuals are morally accountable for actions whose outcomes and even natures they did not intend, if those outcomes and natures were foreseeable. Some philosophers have broadened these points by appealing to a notion of moral luck. See Thomas Nagel, 'Moral Luck' (1976), in *Mortal Questions* (New York: Cambridge University Press, 1979). I think that more exact analysis shows that the phenomena that Nagel discusses are not as destructive of our notions of agency and moral evaluation as Nagel advertises. To avoid relying on any particular view of moral luck, I center on full, unmitigated accountability, and on the applicability and application of moral norms and norms of critical reason to the *individual's contribution* to accountability. I believe that in a sense the individual's contribution is what these norms are *all* about. But the notion of the *individual's contribution* must be understood, and applied, in a nuanced way to make out this "sense". In any case, I stand by the view that a competence for self-understanding of antecedent cognitive and motivating psychological elements grounds full, unmitigated individual accountability, and grounds normative evaluation of the *individual's contribution* to his or her accountability for actions, under the relevant norms.

The ground for buck-stopping status of the *psychological antecedents* of action is that an action's nature or consequences may not conform to the individual's motivations, and may not be foreseeable by the individual. The individual's accountability depends on what he or she intends and is in a position to foresee. The specific sorts of accountability that are the focus of the norms of morality and critical reason depend on those aspects of an individual's point of view that are, at some point, accessible to self-understanding.

Self and hence *person* are kinds that are constitutively determined by capacities for self-understanding and for self-applying certain norms. Aspects of a self or person marked by these starting points are *the core self*. Core selves are partly individuated by their rationally accessible points of view. I want to understand the nature and normative implications of the *self-understanding*—distinguished from the self-application of norms—that is constitutive of core selves.

IV

Let us return to the Kant/Locke conception of a person or self as constitutively having *diachronic* powers of self-consciousness. Kant gives little explicit motivation for his conception (see note 1). Locke provides a motivation that claims too much.

In the previous lecture, I discussed meta-psychological autobiographical episodic memory from the inside as a prime instance of a diachronic self-consciousness constitutive to being a self. I want to explain how this type of memory and an analogous type of anticipation figure in competencies constitutive to selves. I discuss three ways in which the self-consciousness constitutive to selves—and to being subject to moral norms and norms of critical reason—has such diachronic implications: in inference, in practical decision making, and in dialectical aspects of critical reasoning and moral thinking.

The first way lies in inference. I make two assumptions. First, I assume that selves must be capable of propositional inference. Second, I assume that they must be able to think consciously *de re* of some of their propositional inferential events, as their own. The apperceptive rationally accessible point of view must reflexively apply *de re* to inferential events in it. Selves not only must be able to be conscious of *steps*, the premises and inferred contents, in inferences and be able to understand steps as reasons for conclusions. They must also be able to be conscious *de re* of events that make up the inference.

The first assumption is straightforward. As critical reasoners, selves must be able to make rational propositional inferences. The second assumption needs elaboration. Its requirement of consciousness is not what needs it. Being able consciously to carry out a self's constitutive representational functions helps define selfhood. The requirement that one think of inferences as one's own is also straightforward. What needs elaboration is the requirement that selves be able to think *de re* of their inferential events. Two considerations underlie this requirement.

One concerns the role of inference in the buck-stopping status of some of a critical reasoner's states. For any states in a critical reasoner to have such status, some inferences must have it. For being rationally accountable for state instances in one's point of view requires ability to support them by rational inference. Inferences, as token events, must be among the buck-stopping elements in a critically rational point of view. As a consequence of a point I made earlier about induction, if an individual could only infer inductively to what his/her inferences are, inferences could never be the starting point for the accountability of a critical reasoner. So one's access to one's inferences must not always be by inductive inference. Inductive inference is the only *inferential* access one *could* have to one's inference events. Moreover, the access must be through canonical—not purely descriptive—representations of the inferences: the access must be to the inferences on their own terms. Purely descriptive access would not allow the inferences, and their premises, to be starting points of accountability. A noninferential, not-purely-descriptive access is *de re*.

The other consideration behind the requirement of *de re* access to one's inferences concerns the role of such access in having a minimal competence with the concept of reason—a requirement on being a critical reasoner. Minimal competence requires understanding implementation of reasons as such—understanding being moved by reasons in particular cases—in the practice of one's own thinking.[61] If one could not implement reasons as such with respect to state instances that one has in mind—in the practice of one's own thinking—one could not understand implementation of reasons as such. If one lacked *de re* access to one's inferences—and knew them, if at all, only purely descriptively or by inductive inference—they would not be *part* of one's own critical thinking: one could not have them in mind. They would at most be topics of appraisal: like unconscious states that one knows one has, but cannot sustain or change in current critical reasoning. Implementation of reasons as such must be *de re*. Implementation requires *de re* access. If an individual knew only in a non-*de re* manner that he or she had engaged in an inference, the individual still could know whether and why he or she came to a conclusion and perhaps could appraise the inference in some way.[62] But the knowledge would not allow implementation of reason appraisals, any more than one could implement reason appraisals in

[61] I have argued this point at some length in 'Reason and the First Person', 250ff. I think that at least the first, and possibly the second, of the two considerations that I am expounding can be modified to show that having *de re* access to one's inferences is necessary to engaging in the minimal reasoning necessary to being morally responsible.

[62] I think that one cannot engage in reason appraisal if one cannot implement reasons as such—be moved by reasons as such to change or sustain psychological elements in one's own thinking. Cf. 'Reason and the First Person'. Both appraisal and implementation are necessary to have the concept reason. It should be noted that the argument of this paragraph could dispense with this premise. I have used it because it fits with the earlier discussion. The minimal point is that critical reasoning requires that one reflexively hold a state instance in mind as one deliberates with it, evaluates it, and determines according to reason evaluation whether to sustain it or change it. Doing so requires *de re* access to the state instance.

others' thinking. The inference that the individual would know about, the topic of the knowledge, would not be an integral part *in* any critical reasoning. The inference could not be held in mind. So in the absence of *de re* access to his or her own inferences, an individual could neither implement reasons as such with respect to inferences in the practice of his or her own critical thinking, nor minimally understand implementation of reasons with respect to inferences—being moved by them as such. Selves are critical reasoners. Critical reasoners must understand implementation of reasons as such in their own inferences. So to be a critical reasoner, hence to be a self, one must have *de re* access to one's own inferences.

The two assumptions that I have just motivated, in two ways, can be used to show that selves must have meta-psychological autobiographical episodic memory from the inside. The first assumption of a capacity for inference entails a capacity for memory. Inference is not a point event. It requires some representational retention. Retention of mode and content through an inference is an instance of purely preservative memory.[63] Purely preservative memory is not autobiographical episodic memory from the inside. But the requirements on being a self in the second assumption entail that having more than purely preservative memory is required for the relevant inferential capacities.

The second assumption, with its requirement of a capacity to think consciously *de re* of inferential events as one's own, immediately entails that the relevant inferential capacities carry a competence with *meta-psychological autobiographical* memory. *Episodic* memory must be long term, conscious, and of occurrences, as they were at a given past time. It is a form of experiential memory.[64] Experiential memory is *de re* memory of some particular. The requirement that there be a capacity for *conscious* memory is embedded in the second assumption. The requirement that there be an ability to remember inferential processes *de re* entails that the memory is experiential. It is *de re* of events in the inferential process.

What of the requirement of long-term memory? Long-term memory is a technical concept. It contrasts with short-term memory—which is, colloquially, *very* short-term. The various types of short-term memory last a few seconds, or less, and have automatic decay times. The memory required of a self in inference—in moral thinking or critical reasoning—is long term in this sense. At least, a being that had a reasonable lifespan and lacked a natural competence to hold together propositional inferences—or practical decisions or deliberative dialectic—longer than a couple of seconds would be rationally deficient in a way that would fall short of having the powers essential to being a person or self. (See note 102 below.)

[63] *Purely preservative memory* is retention of mode and content from an earlier psychological state, retention that does not introduce new subject matter and is not part of a justification for the conclusion of an inference. For discussion of purely preservative memory, see my 'Content Preservation', *The Philosophical Review* 102:4 (October 1993), 457–488; 'Memory and Persons'.

[64] See 'Memory and Persons'. Recall that experiential memory need not be warranted empirically. It is not a type of inner *sense*. See note 37.

Requiring that a capacity for meta-psychological autobiographical episodic memory in inference be *from the inside* entails requiring the memory to be capable of veridically preserving the perspective of the inference. This requirement follows from the second assumption. The relevant sort of *de re* thinking about one's inferential processes involves correctly remembering the mode and content of particular inferential steps through purely preservative memory. When meta-psychological autobiographical episodic memory relies on purely preservative memory in this way, the former type of memory preserves the perspective of the inference and represents it as it was when it occurred.

The two assumptions show that meta-psychological autobiographical episodic memory from the inside, or some short-term analog, is constitutive to selves. Such memory is clearly a type of consciousness of one's past self, plausibly in the sense of Kant and Locke.

I turn more briefly to a second way in which diachronic capacities are constitutive of being a self. I believe that both the special type of memory that I have been discussing and an analogous capacity for anticipation of the future are constitutive to selves, moral beings, and critical reasoners by virtue of their constitutively having *practical* reasoning. Practical reasoning has both forward-looking and backward-looking diachronic dimensions. Intentions have the representational function of being fulfilled by the intender. When they are fulfilled through occurrent decisions to act, and by actions themselves, there is a match with the antecedent intention. Self-conscious selves can recognize these points. Looking forward from an intention, an individual is aware of his or her being the prospective executor of the intention. The individual can think anticipatorily of him- or herself, in the first-person way, as preserving the intention and as realizing it, through a decision to initiate action and through the action.[65] Remembering backward from a decision to initiate action, or from an action, a self-conscious individual can recognize moving on the antecedent intention, and satisfying or failing to satisfy the intention. These diachronic capacities are integral to the self-understanding that is constitutive of being a self and of being subject to moral norms and norms of critical practical reason. They are consciousness of oneself as one will be in the future, as well as consciousness of oneself as one was in the past.

An argument similar to the one given for propositional inference shows that requirements on having the reflexive powers with regard to decision making that are required of selves requires meta-psychological autobiographical episodic memory from the inside—and an analogous *de re* capacity to anticipate particular representational events in one's future.

[65] Framing possibilities for action is an aspect of selves' making choices. Selves have a probably constitutive capacity to think of themselves not only diachronically, but modally. Locke writes of selves as being able to be conscious of themselves in different times *and* places. Cf. *An Essay Concerning Human Understanding*, 2:27: 9. He may have had in mind the relatively primitive freedom of imagination to think of different spatial possibilities in action.

A third source of the requirement that selves have self-conscious diachronic capacities lies in the dialectical reasoning integral to deliberation. A condition on being subject to norms of critical reason and morality is to be capable of dialectical self-criticism.[66] A person or self must be able to stand behind a belief or decision after reflecting on it, to answer self-questioning, and to effect change of mind after self-criticism. Such dialectic requires an ability to think *de re* of one's thoughts, while being committed to them and while criticizing past or prospective commitments. In critically evaluating a present commitment, one must be able to conceive oneself as possibly giving it up. In changing one's mind, one must be able to conceive oneself as having given up an earlier commitment, knowing why. Recognizing actual or possible differences in one's point of view, retrospectively or prospectively, is part of being a *critical* reasoner. Diachronic forms of self-consciousness are part of recognizing what accountability is—a capacity to support or change one's position through rethinking or new information. Such capacities to recognize differences in one's own point of view, as they unfold in dialectical reasoning, are constitutive to being a self.

The diachronic use of self-understanding that is partly constitutive of apperceptive rationally accessible points of view combines *episodic* memory with *purely preservative* memory. Episodic memory gives one diachronic *de re* access, allowing states other than immediately present ones to have buck-stopping status and to be part of the thinking practice that one implements reasons within. Purely preservative memory gives one diachronic access to the relevant psychological elements on their own terms—also a central requirement on their entering into critical reasoning and moral accountability.

To elaborate this latter point: purely preservative memory retains the mode *type* and representational content of psychological state instances. Analogously, for anticipation of the mode and representational content of one's future point of view. Mode and content are the aspects of psychological state instances that are relevant to application of norms of critical reason and morality. If diachronic self-consciousness could not preserve the content and inferential order of past psychological events, one could not review one's inferences and commitments on their own terms. If diachronic self-consciousness could not anticipate the mode and content of instances of one's future points of view, one could not control one's future acts. Reflective review and anticipatory control are basic functions of a self.

I believe that the foregoing considerations provide a deeper characterization and motivation of the relation between selfhood, or psychological personality, and diachronic self-consciousness than we find in Locke or Kant.

[66] Dialectical reasoning is a special case of agency. So the points in this paragraph specialize the points in the previous paragraph. I distinguish practical agency and agency involved in theoretical reasoning. Dialectical reasoning can be either practical or theoretical. The agency involved in reasoning to intentions or to act-initiating decisions is practical.

V

I conclude by summarizing features of the relevant self-understanding. These features will ground discussion of its epistemic warrant. That topic will dominate the third lecture.

The self-understanding must be *meta-psychological*. The norms that I have discussed demand understanding the norms and applying them to one's point of view. To understand application of the norms, it is not enough to have *competence understanding*—the capacity to *use* concepts. One must be able to think about propositional attitudes, and their representational contents, as such. This is meta-psychological understanding.

Second, the meta-representational understanding must be *phenomenally conscious* and *rational-access conscious*. There must be a what-it-is-like quality to the understanding. And the lower-level, "understood" aspects of the point of view must be occurrently accessed by the individual's rational powers. The requirement that a self be phenomenally conscious, some of the time, is so basic that it is hard to motivate. Except in cases of understanding sensations and other states constitutively associated with consciousness, phenomenal consciousness is less important epistemically than in marking off the psychological kind, *self*, that underlies our deepest normative valuations. Those states that are constitutively associated with phenomenal consciousness (pleasure, pain) ground many primary moral valuations. Beings that lack a natural capacity for phenomenal consciousness do not have the moral claim on us that phenomenally conscious beings do. The requirement that a self be able to make accessible to conscious rational control some lower-level aspects of its point of view is just the product of combining the requirement of phenomenal consciousness with the role of self-understanding in applying the norms to one's own case.

Third, the understanding must exercise *self-conscious* meta-representation. A self must have a first-person self-concept and must apply it consciously. It must recognize aspects of its point of view as its own. In thinking I believe that Brahms is greater than Chabrier, I not only refer to myself with I; I take the belief that I represent to be mine. This capacity to represent elements of one's point of view as one's own is, of course, basic to the self-evaluation that grounds being subject to moral norms and norms of critical reason.

An aspect of this self-consciousness is that both the lower-level psychological element and the meta-representational element that constitutes the self-understanding are reflexively *understood*. One is accountable for the meta-representational self-understanding as well as for the understanding's subject matter, the lower-level psychological element.

Fourth, the understanding must include past and future as well as present elements in one's point of view. I have highlighted this constitutive requirement because I believe that it has been less well understood in the philosophical tradition than some of the others.

Fifth, the self-understanding must be *systemic*. This requirement underlies the fourth one. The self-understanding represents not just psychological states and their representational contents. It represents relations among the attitudes and the contents. Reason is essentially relational. An understanding that applies the concept reason must specify relevant relations. Moral evaluation of acts must relate them to beliefs and intentions. Even where moral norms evaluate attitudes, they must do so *relative* to information accessible to the individual.

Sixth, the self-understanding must be *specific*—from the inside. To ground norms of morality and critical reason, the self-understanding must specify psychological states in ways that preserve the point of view to be evaluated. The specifications must enable one to know what the point of view is on its own terms. Since the norms concern the perspective of the individual, an understanding that did not preserve the mode and content of the evaluated psychological elements would be too detached to render the norms applicable. Of course, much is forgotten. But an individual's having a *capacity* to preserve the mode and representational content of evaluated attitudes is constitutive to applicability of the norms.

Seventh, the relevant self-understanding must be *immediate*, non-inferential, in the apperceptive rationally accessible point of view.[67] Access to self-understood psychological elements is not via individual-level inductive inference. Psychological elements immediately accessible to self-understanding are starting points for expression of an individual's moral being or critically reasonable self. They have buck-stopping status for evaluation under the relevant norms. Psychological elements that are not immediately accessible cannot ground applicability of the norms of morality and critical reason. The starting points of, and transitions in, some inference must be capable of non-inferential self-understanding.

There is an existence proof of individual-level non-inferential elements of self-understanding. It is provided by Descartes' pure and impure *cogito* judgments. The structure of these judgments shows clearly that they are not inferential. When I judge I am thinking, or I [hereby] judge that Brahms is greater than Chabrier, or I intend to take the metro, the judgment is reflexive in a way that precludes individual-level inference. The judgment about the psychological state has, as a constituent part, the psychological state instance that the judgment is about. Judgment and subject matter are part of a single thought.

Although *cogito* cases do not involve any individual-level inference, unconscious psychological causal processes connect some relevant types of self-understanding with their psychological subject matters. For example, all self-understanding that includes meta-psychological autobiographical episodic memory relies on sub-individual psychological transitions. Whether these are propositional inferences is an empirical matter. I think that they probably are not, but I rest nothing on this belief.

[67] One can work out inductively things about one's point of view from inside it, as long as one starts one's inferences from a self-understanding that is immediate and from the inside. Wisdom about oneself lies in that direction.

What is important is that these memories are not results of *individual-level* inferences. *Within the apperceptive rationally accessible point of view*, some of them are immediate. Any "easy" inferences to self-understood states in a rationally accessible apperceptive point of view must be open to non-inferential self-understanding of the inferences' starting points and transitions. If an individual comes inductively to a view about a state that is not immediately or easily accessible, and relies entirely on the induction for self-understanding, the individual approaches one aspect of his or her point of view from the perspective of another. The understanding is not from the inside. For it does not preserve that point of view.

Psychological elements can have buck-stopping status only if they can be self-understood from the inside. They must be capable of being self-understood *in* the relevant psychological acts, states, or processes. These points apply to autobiographical memories. To self-understand past states in ways that give those past states buck-stopping status, some of the memories must be from the inside. Hence those states cannot be understood entirely on the basis of induction, within the apperceptive rationally accessible point of view.

An eighth feature of the self-understanding is its *de re particularity*. This particularity has several aspects. Recall that *de re* representation is representation that picks out a particular via a non-inferential, not purely attributional or descriptive competence.

The understanding is *de re* in four ways. First, the application of the self-concept is *de re*. Application of the self-concept is not purely attributional or descriptive. It is effected through a rule that infallibly effects reference through the mere occurrent *use* of the concept.

Second, one's understanding of the *representational content* of one's psychological states is *de re*. Self-understanding uses canonical representation of representational content. Such representation uses the representational content in the very representation of it. When one understands one's belief that Federer will win a match, canonical representation of the belief 's representational content (that Federer will win a match) refers to the representational content, and also uses that representational content. Such canonical representation is *de re*.

Third, self-understanding of elements in one's point of view specifies them by applying a temporal index. This indexing is *de re*. One singles out the time non-descriptively and indexically, roughly as *now* or *this time*. In self-understanding that uses episodic memory, there are two *de re* temporal indexes: one for time of the self-understanding, the other for the past time of the understood psychological element.

Fourth, the self-understanding is *de re* with respect to particular state instances, acts, or occurrences. One understands instances of psychological types, not just repeatables. In singling out instances, one uses concepts—the canonical name of the representational content, the self-concept, temporal indexicals, and concepts for psychological modes (like believes, intends, perceives as of). But the use is not purely conceptual. One also occurrently *applies* the concepts to yield *de re* understanding of psychological particulars. In thinking I am wondering if Israel

will make peace, I refer to an occurrence of wondering. In thinking, non-inferentially, I believe that Rembrandt is deep, I represent a standing instance of the belief.

It is constitutive to the role of self-understanding in grounding applicability of these norms, and in constraining their application, that it apply *reflexively* to the particular point of view and the particular psychological states of the understander, not just the same type of point of view. The individual must understand applicability of the norms to elements in his or her *own* point of view, *understood as such*.

I have been discussing *representational* aspects of the self-understanding that helps ground selfhood, ground the applicability of moral norms and norms of critical reason, and constrain application of these norms. It is meta-representational; conscious; self-conscious; diachronic as well as synchronic; systemic; specific; immediate, or non-inferential, at the proprietary individual level of the psychology; and, in multiple respects, *de re*. Next lecture, I explore the *epistemic* status of the self-understanding—connecting representational structure and representational competence marked by that structure with their epistemic credentials.

9 *Self and Self-Understanding*

Lecture III: Self-Understanding

I have taken selves to be constitutively self-conscious critical reasoners, and arguably constitutively subject to moral norms. I isolated a type of self-understanding required for applicability of these norms. The norms require an individual to be able to understand some of his or her psychological states from inside having those states.

Norms of critical reason and morality hinge on elements in an apperceptive rationally accessible point of view. Such a point of view is that part of a point of view whose elements are immediately or easily, presently, consciously accessible to self-attribution. Self-understanding helps give such elements *buck-stopping status* with respect to the two norms. Having buck-stopping status is having most fundamental status in explanations of the applicability and applications of a norm.

I

I want to understand the *epistemic warrant* for this self-understanding. To explain any epistemic warrant, one must explain what makes a warranted state an objectively good route to truth, given an individual's background information and competencies. Being a good route implies reliable veridicality in normal conditions. Normal conditions are conditions by reference to which the state's representational nature is constitutively determined.[68]

A warrant must also "make sense of" or rationalize the individual's point of view. This condition is hard to make precise. It is grounded in the idea that

[68] For an argument that normal conditions, in this sense, are the relevant conditions in which the reliability of the underlying warrant is to be considered, see my 'Perceptual Entitlement'. This account shows how the condition of reliability, and warrant for one's beliefs, can remain in place even in conditions—such as brain-in-vat conditions—in which one's beliefs are highly unreliable. Although *de facto* reliability—for example, in abnormal conditions—is not constitutive of being warranted in the empirical case, *de facto* reliability in one's access to some of one's own mental states *is* constitutive of being a self. Individuals that lose all reliability in judgments about any of their mental states, say through pathology, lose their selfhood.

warrant is a norm for psychological competence. There must be something *about the nature of the individual's capacities or point of view* that helps explain the individual's psychological states' being a good route to truth. Brute, "accidental" reliability does not suffice for warrant.

It would be too much to require an individual to be able to explain the warrant. Animals, children, and even adult persons can be warranted in perceptual or other beliefs, and yet be unable to explain why. On the other hand, correct explanation, even if the individual cannot give one, must center on the individual's point of view. For what is being made sense of or rationalized, under epistemic norms, is the individual's competence and perspective.

Epistemic warrant for a belief is sufficient for knowledge, if the belief is true and if its relation to truth is in certain ways not abnormal or adventitious. Warrant concerns the contribution to goodness of route to truth by the *individual's* competencies. Failures of knowledge do not reflect specifically on the individual or the individual's psychology if the individual's states are warranted. In this way, warranted but false beliefs are like good acts with bad consequences. I call a false but well-functioning and warranted propositional attitude, warranted in way *W*, an attitude that incurs *brute error* relative to warrant *W*.[69]

Many epistemic warrants are compatible with brute error. Nearly all *empirical* warrants are compatible.[70] Perceptual beliefs can be warranted but mistaken because of an illusion that one lacks ground to reject. Reasonable inductive empirical explanations can be mistaken. Even most empirical *knowledge* rests on warrants that are compatible with error.

The warrants for self-understanding and self-knowledge that we are exploring show a different pattern. The warrants yield *immunity to brute error*: if the psychological state is warranted, it is veridical. Before developing this point, I shall discuss it intuitively.

Many psychological states are not easily accessible to self-knowledge. Then a belief 's being warranted is compatible with brute error. Knowing one's deeper motivations, or character, or the unconscious states and transitions that populate the mind, is hard. Error can derive from the sheer difficulty of the psychological subject matter. Often, mistaken self-ascription is unwarranted. But warranted error can certainly also occur.

By contrast, in many types of cases, if a mature person is wrong about his or her beliefs, intentions, percepts, or sensations, it is common to infer that there is something wrong with the person. Errors reflect on the use or health of the person's cognitive equipment. If a person were wrong about his or her *belief* about whether he or she has a sibling, it would be normal to infer that the person

[69] I introduce the notion of brute error in 'Individualism and Self-Knowledge', *The Journal of Philosophy* 85:11 (November 1988), 649–663; cf. also 'Perceptual Entitlement', section II.

[70] Exceptions are knowledge of one's sensations and perceptions. These are empirical because their warrants depend for their force on *having* of the sensation or perception. I later discuss why at least paradigm cases of these sorts of knowledge are immune to brute error.

was self-deceived, or had allowed emotion to cloud judgment, or was afflicted with some pathology. It is easier to imagine cases in which one has a warranted mistaken belief about whether one *has* a sibling. I am evoking the oddity of a warranted but mistaken belief about whether one *believes* that one has a sibling. Here error commonly reflects on how one formed one's meta-belief. Often, a mistake entails malfunction or other non-standard belief formation, which undermines warrant.

These issues are delicate. The *natures* of *most* psychological states that are accessible to relevant self-understanding do *not* guarantee that they are accessible.[71] Being a belief that one has a sibling does *not* insure susceptibility to self-understanding that is immune to brute error. But I shall argue that, for selves, there are *necessarily some* cases of immunity to brute error, and that such cases ground applicability of norms of critical reason and morality.

I cannot emphasize too strongly that except in *cogito* cases, self-attributions that are warranted in the relevant way *cannot* be determined by the contents and modes of the attributed states. Evidence as to whether one has a sibling could be mixed. One could be unsure of the nature of one's own reaction to the evidence: do I believe that I have a sibling? Using behavioral evidence about oneself, one could be warranted in concluding that one does believe one has a sibling. But one could lack that belief. One could be puzzled, or have a leaning that falls short of belief. Then, the warrant for the mistaken belief that one believes that one has a sibling would not be the warrant that I am reflecting upon. The warrant would not be immune to brute error. Except in *cogito* cases, one cannot generalize apriori about *which* of one's self-attributions can be warranted in the special way. What is apriori determinable, I think, is that *some* psychological states in selves are susceptible to a type of self-understanding that is immune to brute error, relative to the relevant warrant.

Intuitively, we sometimes do understand our own psychological states in a specially warranted way. This intuitive point is supported by theoretical considerations. Let us reflect on the relation between immunity to brute error for the relevant self-understanding and the buck-stopping status of self-understood psychological states in critical reason and morality. Immunity of the self-understanding to brute error is constitutive to having an apperceptive core self. A type of self-understanding that is *liable* to brute error cannot, I think, ground the buck-stopping status of attitudes in applications of these norms. If an individual's self-understanding of a psychological element can be warranted but mistaken, the self-understanding is too detached from that element to give that element buck-stopping status. That is, if one can understand a psychological element only in a way that is liable to brute error, that element cannot be a basic starting point for assessing the individual's contribution to full unmitigated accountability under the norms of critical reason and morality.

[71] Exceptions are pure *cogito* cases and some knowledge of conscious sensations.

Being subject to brute error means that in reviewing one's psychological state or event, one is not fully accountable for getting that state or event right. Error is compatible with one's best use of one's competence for self-understanding. Such reviews do not constitute exercise of a self-understood point of view that one has cognitive control over. They are not exercises of a cognitively unified perspective. The norms of critical reason and morality take the basis for their evaluations to be psychological states that are understood from inside the perspective of those states—preserving that perspective meta-cognitively in the individual's self-review. The norms require the individual to be able to understand what he or she does or undergoes *in* the doing or undergoing, from the inside. If one could not take up the same perspective that those acts or endurings had from the inside—incorporating it into the meta-representational review, by preserving it or making it part of a single meta-representational perspective—one would always be in the position of a privileged outsider. Psychological elements self-understood in ways that allow for brute error lack buck-stopping status in persons and selves.

The idea is that, ultimately as a critical reasoner and as a moral being, an individual's self-understanding must inform, or be capable of informing, some of the individual's cognitions, reasoning, decisions, and other states central to being a self. The self-understanding must inform them in a way that is part of their exercise or part of their being experienced. Working out an inductive, empirical understanding of one's reasoning and decisions, or relying on some other sort of understanding of them that is subject to brute error, is not the kind of self-understanding that forms a basic starting point for critical reasoning or morally responsible decisions. The self-understanding must be from the inside—from a meta-perspective that preserves and is informed by—the psychological states and acts that one is most basically and unmitigatedly responsible for as a critical reasoner and moral being. One must be able to self-understand from the inside those states, acts, and events that are fundamental to being a self.[72]

There are certainly types of self-understanding that rest on warrants that leave one vulnerable to brute error. Immunity to brute error is not an honorific status. It is not invulnerability to all error. It is a normative standing that hinges on possible sources of error, relative to epistemic norms. The deepest wisdom in self-understanding—a result of learning about oneself—*is* vulnerable to brute error. I claim only that the self-understanding required by norms of morality and

[72] I make a version of this point in 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 96 (1996), 91–116; cf. 109ff. See also the discussion of (CR) below.

One can *get to* a veridical self-understanding that replicates the perspective of a psychological state by inductive inference. If the self-understanding relies on the inference, it would still be subject to brute error, as all induction is. But it would not have buck-stopping status. It could not ground the applicability of norms of critical reason and morality. Only the starting points of the inference could have buck-stopping status, assuming that they are susceptible to the right sort of self-understanding. The starting points of the inference could not have buck-stopping status and be subject to brute error, if they are to be *from inside* the perspective of the understood psychological states, in the relevant sense.

critical reason—and required for there to *be* selves—is immune to brute error: being warranted is incompatible with error. This is a less comprehensive, less insightful self-understanding than wisdom about oneself. But it is the sort that is constitutive to being a self, and to the buck-stopping status of self-understood states in the applicability of norms of critical reason and morality. From here on, I *capitalize* ('Self-Understanding', 'Self-Knowledge') to indicate the self-understanding and self-knowledge that is my focus. *Sans* capitalization, the terms express generic notions that apply to all types.

II

In the last quarter century, self-knowledge has come in for a lot of discussion. I will not canvass the literature. I ignore views that assimilate Self-Knowledge to inner perceptual belief or to empirical hypotheses. I and others have criticized such views elsewhere.[73] I also ignore views that deflate Self-Knowledge, holding that it is trivial or non-substantial. I will, however, say a little to distinguish my view from others.

The psychological competencies that engender Self-Knowledge are not in general infallible. Pure *cogito* thoughts *are* infallible. If one thinks I hereby in this very act think [entertain] the thought that *Figaro* is the greatest opera, that thought about my thinking cannot be mistaken.[74] But pure *cogito* cases comprise only a tiny portion of Self-Knowledge. Immunity to brute error in Self-Knowledge is not immunity to all error. Non-inferential self-attributions of the relevant types can be mistaken because of unwarranted interference or pathology. The epistemic specialness takes subtler forms than infallibility.

The competencies that yield Self-Knowledge are also not to be understood in terms of self-intimation. I doubt that *any* type of psychological state is such that if an individual is in it, it is impossible for the individual not to believe that he or she is in it.

Some traditional accounts of self-knowledge explicate its epistemic specialness in terms of the nature of the subject matter. Conscious mental elements are supposed to render judgments about them epistemically special because of the elements' very nature. Thoughts were said to be transparent, self-evident,

[73] Sydney Shoemaker, 'Self-Reference and Self-Awareness', *The Journal of Philosophy* 65:19 (Oct. 3, 1968), 555–567; Shoemaker, *The First-Person Perspective and Other Essays* (New York: Cambridge University Press, 1996); my 'Individualism and Self-Knowledge' and 'Our Entitlement to Self-Knowledge'; Richard Moran, *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton: Princeton University Press, 2001); Akeel Bilgrami, *Self-Knowledge and Resentment* (Cambridge University Press: Harvard, Mass., 2006), chapter 1; and Lucy O'Brien, *Self-Knowing Agents* (New York: Oxford University Press, 2007), chapter 3.

[74] See my 'Individualism and Self-Knowledge'; Andreas Kemmerling, 'Eine reflexive Deutung des Cogito', in Konrad Cramer et al. (eds.), *Theorie der Subjektivität* (Frankfurt: Suhrkamp, 1987); Kemmerling, *Ideen des Ichs: Studien zu Descartes' Philosophie* (Frankfurt: Suhrkamp, 1996). The original source is, of course, Descartes, *Meditations on First Philosophy*, II.

unmistakable, or self-intimating by their very nature—as if they glowed in the dark. Such accounts are neither explanatory nor accurate about the psychological or epistemic phenomenon.

Some recent accounts have made claims—rhetorically different, but I think ultimately similar—that it is of the nature of 'genuine', 'full-blown' intentionality that mental states be known or normally knowable by any individual who has them.[75] Animals and very young children are attributed a lesser type of intentionality, or none at all. I think that such views are hopeless. Their conceptions of "genuineness" in intentionality lack serious defense. The accounts do not accord with scientific explanations that attribute perceptual, intermodal, and propositional intentional (representational) states to animals and very young children. Many or all of these individuals lack self-knowledge. They have intentional, representational states in the most literal sense of these terms. No argument has shown that their psychological states are any less genuinely representational than those of human adults. There is no difference in kind or degree in their *representationality*. Self-Knowledge, and self-knowledge, cannot be explained in terms of the natures of first-order representational states. There is nothing in their natures that insures that they are self-known or self-knowable.

The natural phylogenetic and developmental order is to have representational states before being able to know them. This natural order is the order of language acquisition. Even if higher-order representational concepts are innate in human beings, which I doubt, no one has given good reason to think that there is a constitutive connection between an individual's having representational states, even having propositional attitudes, and that individual's being in a position to know what those states are.

Another claim is that the subject matters of self-knowledge constitute *reasons* for the knowledge.[76] Both lay-people and philosophers think about reasons in disparate ways. I make just a few points here.

Reasons are necessarily representational. They are considerations that are good routes to truth, in the case of theoretical reasons, or good routes to goodness or aptness of acts, in the case of practical reasons. A thunderbolt or a sea otter cannot be a reason. Only some propositional representational content, with some associated mode—for example, judgment—about the thunderbolt or sea otter can be a reason. Nature *grounds* reasons.[77] It provides entities and situations representation of which can form a reason. But reasons are constitutively propositional

[75] Donald Davidson, 'Rational Animals' (1982), 'The Second Person' (1992), and 'The Emergence of Thought' (1997), all in *Subjective, Intersubjective, Objective* (New York: Oxford University Press, 2001); Bilgrami, *Self-Knowledge and Resentment*, cf. pp. 160ff., 178–182, 205.

[76] Christopher Peacocke, 'Conscious Attitudes, Attention, and Self-Knowledge', in Crispin Wright, Barry C. Smith, and Cynthia Macdonald (eds.) *Knowing Our Own Minds: Essays on Self-Knowledge* (New York: Oxford University Press, 1998).

[77] A major source of the confusion lies in not distinguishing support in reasoning and argument—reasons—from support of other kinds, including physical evidence. For a good account of such distinctions, see Robert Audi, *The Architecture of Reason: The Structure and Substance of Rationality* (New York: Oxford University Press, 2001).

and representational. Nature is not a text and is not made up of propositions or reasons. Since reasons are necessarily *representational*, sensations cannot be reasons unless they are representational. I believe that some sensations are not representational. In such cases, only representation that one has the sensation can be a reason.

Reasons are necessarily *propositional*. Reasons are potential steps in arguments and explanations that, in effect, show why one should believe or intend something.[78] They answer potential why questions. As a matter of logical grammar, non-propositional—and non-representational—entities cannot play these roles. Non-propositional entities, such as perceptions, can contribute to an attitude's being warranted. But reasons are necessarily propositional contents taken with their modes. They are basic units in reasoning. Some Self-Knowledge is of sensations and perceptions, which are not propositional. These entities *could not* be reasons for Self-Knowledge. But it can be shown independently of this point that what one knows in Self-Knowledge is not a reason that supports the knowledge. Let us focus on knowledge of one's own propositional attitudes.

Are first-order beliefs reasons to judge that one has those beliefs? They cannot be. A reason, theoretical or practical, must be about the subject matter of the attitude for which it is a reason, or about some relevantly related subject matter. The reason's representational content must bear on, and in some way count in favor of, the truth or goodness of the conclusion for which it is a reason. First-order attitudes are not about the subject matter of second-order attitudes. They commonly bear no relevant relation to it. They *constitute* the subject matter, but they are not reasons about the subject matter that support the truth or goodness of the higher-order attitudes. If one's first-order attitudes were used as reasons for judgments that one had them, one's reasoning would consist of *non sequiturs*.

Most Self-Knowledge is not supported by reasons. The few exceptions are self-evident *cogito*-like cases. Not being supported by reasons does not make Self-Knowledge any less epistemically warranted or substantial. I shall elaborate this point shortly.

An idea with some currency is to explicate Self-Knowledge by reference to knowledge of one's intentional physical agency.[79] I myself emphasize Self-Knowledge's role in the applicability of norms of morality and critical reason.

[78] I say 'in effect' because the reason need not respond to an actual question, and because a reason is commonly not a meta-level consideration that makes reference to shoulds, beliefs, intentions, or representational contents. Reasons can be about object-level matters. They are reasons by virtue of their potential use in supporting and explaining the "why" of conclusions.

[79] See Moran, *Authority and Engagement*; Matthew Soteriou, 'Mental Action and the Epistemology of Mind', *Noûs* 39:1 (March 2005), 83–105; Bilgrami, *Self-Knowledge on Resentment*; O'Brien, *Self-Knowing Agents*; Peacocke, 'Mental Action and Self-Awareness (I)', in Jonathan D. Cohen and Brian P. McLaughlin (eds.), *Contemporary Debates in the Philosophy of Mind* (Malden, Mass: Blackwell, 2007), 358–376; Peacocke 'Mental Action and Self-Awareness (II): Epistemology', in O'Brien and Soteriou (eds.), *Mental Actions* (New York: Oxford University Press, 2009), 192–214. Some of these approaches are not strictly incompatible with what I say here, but all highlight agency in self-knowledge more than I do.

These norms centrally concern agency. It is certainly true that *some* of what is known in Self-Knowledge comprises mental acts. Persons and selves—the subject matters of Self-Knowledge—are constitutively agents.

The approaches that I have in mind, however, highlight agency in a more specific way. They model Self-Knowledge—or some types of Self-Knowledge—on an individual's knowledge of his own intentional *bodily* action. Knowing one's thoughts and attitudes is taken to be structurally and epistemically like knowing that one is raising one's arm.

I cite two grounds for not following this line. One is that doing so cannot account for Self-Knowledge's buck-stopping role in morality and critical reason. It is not one's bodily actions, but their psychological antecedents that ground these norms. Assessing full, unmitigated accountability for bodily actions depends on assessing their conforming to potentially Self-Understood cognitions, intentions, decisions, and intentional initiations that produce them. A corollary is that warrants for beliefs about one's bodily actions do not yield immunity to brute error.[80] The warrants I seek *do* yield immunity.

A second ground for not modeling Self-Knowledge on knowledge of bodily action is that a lot of Self-Knowledge is not of acts. Selves *are* constitutively agents. Thinking morally or critically requires a self-consciousness and control that insure that the thinking is agential. Critical reasoning requires judgment, an act. Persons and selves constitutively make decisions in the light of values, again acts. All these points bear on constitutive aspects of being a person or self. But persons and selves, *we*, are not *just* agents. We not only act. We undergo and endure. Knowledge of passive elements in one's mind is virtually as important to norms of morality and critical reason as knowledge of active elements.

Some Self-Knowledge centers on sensations and perceptions that we do not bring upon ourselves. Some Self-Known propositional thoughts simply occur to one. Most emotions are not acts. Even many propositional commitments—all standing commitments (standing beliefs or fears) and many occurrent commitments (many occurrent perceptual beliefs)—are not acts. Many do not even derive from acts. Many are acquired through perception and are first stored automatically.[81] Sometimes in Self-Knowledge one reaffirms such states in

[80] See discussion in Lecture II, section III of the buck-stopping status of the psychological antecedents of bodily actions. The immunity to brute error of Self-Knowledge of the psychological antecedents derives from the same fact that gives the psychological antecedents buck-stopping status. Knowing the psychological antecedents depends only on good use of the competence for Self-Understanding. Knowing one's actions through knowing the psychological antecedents is subject to contingencies involved in realizing one's intentions. One can be fully warranted in one's belief about what one's action is, utilizing one's warranted belief about what one is deciding to do. But one can still be mistaken about the action because of occasion-dependent, post-initiation distortion of the bodily action. Then, the individual's contribution to accountability lies fundamentally in the Self-Understood, or Self-Understandable, psychological antecedents. Accountability for the bodily action commonly is mitigated. Mitigation tracks the individual's brute error about what the action is.

[81] Many philosophers emphasize consciousness as a constitutive feature of intentionality or even knowledge. There are, however, many unconscious representational and knowledgeable states. If consciousness makes a type of state more reliable, it is epistemically relevant. This is an empirical

judgments—which *are* acts. But often in Self-Knowledge one simply forms a belief about what beliefs one has long had. The truth and warrant involved in such meta-beliefs do not hinge on reaffirming the standing commitments that did not arise through any agency. Such non-agential states and occurrences are susceptible to Self-Knowledge.

Knowing one's pains, fears, and other aspects of vulnerability is crucial to one's moral status. Knowing one's perceptual perspective and other passive representational states is crucial to critical reasoning. Self-Understood passive elements in a psychology loom large among elements with buck-stopping status for both morality and critical reason. Indeed, Self-Understanding itself can be passive—the product of the triggering of standing competencies.

III

Let us turn from paths not taken to our main route. I noted earlier that most Self-Understanding and Self-Knowledge are not based on reasons, but are nonetheless warranted. I distinguish two types of epistemic warrant. A warrant that is not based on a reason is an *entitlement*. Such warrants need not be accessible, even in principle, to the individual whose thoughts or attitudes are warranted. A warrant that is based on a reason is a *justification*.

Apart from *cogito* cases, I think that the special epistemic warrant underlying Self-Knowledge is an *entitlement*, not a justification. The warranted individual need not have a reason for the warranted belief. On the other hand, the entitlement requires more from an individual than many other entitlements do. A young child can be entitled to perceptual beliefs, even though it has only the competence understanding required to have the beliefs. To be warranted in Self-Knowledge, an individual must be competent in employing meta-representational attitudes, and must think the content of the attitudes that are the subject matter of the Self-Knowledge, while specifying them in the canonical, that-clause-like way.

I believe that warrants for attitudes that constitute Self-Understanding and Self-Knowledge yield *immunity to brute error*. If an attitude *type* is associated with competencies that could ground a warrant W_{ibe} that yields immunity to brute error, any error in an *instance* of an attitude of that type must be explained in one of three ways. First, the attitude instance may be warranted only in another way W that does allow brute error. Second, the instance may derive from misuse of competencies that undermine warrant W_{ibe}. For example, bias, self-deception, or emotion might derail a normally warranted belief formation. Third, the individual might suffer some pathology that undermines W_{ibe}. A lesion might cut off judgment from normal access to intentions or beliefs. In such cases, an

matter. There is probably unconscious self-knowledge that is analogous to Self-Knowledge in being epistemically "special". Any such self-knowledge is not part of the apperceptive core self.

individual lacks the *relevant* epistemic warrant. Having an epistemic warrant that yields immunity to brute error guarantees *getting things right*. Being relevantly entitled to an attitude about one's psychological states ensures that the attitude is true.

Entitlements to Self-Knowledge are not the only warrants that yield immunity to brute error. I shall reflect on three other classes of cases.

The first class includes three types of entitlement associated with first-order deductive inference. In discussing this class, I distinguish first-order inference from such inference *supplemented* by thinking the inference rules that warrant steps in the inference. Individuals can carry out inferences without having the meta-representational concepts, or the ability to generalize, necessary to think the inference rules that justify their inferential transitions. Even if they do have these abilities, the abilities may not be operative in a good deductive inference. Reasoners can be warranted in their reasoning if the reasoning is governed by—correctly explainable in terms of—following correct rules of inference, even if the reasoner lacks resources to think the meta-representational rules, consciously or unconsciously, or does not rely on such resources in carrying out the inference. The *steps* in the inference can be justifications, reasons, for later *steps*. The warrants for the transitions between steps can, however, be entitlements—warrants without reasons. Reasoners can be entitled to rely on the transitions as valid, even though they cannot, or do not, think the inference rules. Entitlement derives from competence with the logical constants in carrying out the reasoning.

The entitlement to make a deductively valid transition yields immunity to brute error in making the transition. One can make a mistake in such transitions. But if one is entitled to a transition in a deductive inference that relies on purely preservative memory, one cannot make a mistake in the reasoning transition.[82] Being entitled to a transition entails that the transition is free of error.

[82] Equivocation between "twin" concepts in "slow-switching cases" requires that this point be understood with some subtlety. I think that errors in reasoning that depend on reasoners' mixing up "twin" concepts that they have unawares rely on mistaken presumptions of equivalence between the twin concepts. In this respect they are like mistaken presuppositions or even mistaken tacit beliefs. The errors do not result from failures of anaphoric, purely preservative memory, since such memory is not relied upon in the inference. They occur when a reasoner relies on sources—such as substantive memory or perception—that introduce content into later premises in an argument *without* relying on purely anaphoric, preservative relations to the content of earlier premises. Nor do the errors result from failures of rational deductive competence. They involve what Luca Struble calls a brute error in exercising a coordination competence, not an error in a reasoning competence *per se*. Although the presumptions of sameness or equivalence of "twin" concepts are not tacit beliefs, since no state puts the two concepts together into a proposition, the presumptions have much the same function. It is clear that possible errors in inference that hinge on equivocation between "twin" concepts are brute errors. They do not reflect on the well-functioning of *memory* or of *reasoning competence*. So the errors do not have their sources in memory or reasoning competence. I introduced slow-switching cases in 'Individualism and Self-Knowledge'. Paul Boghossian thinks that such cases cause problems for anti-individualism in accounting for self-knowledge and inference. See his 'Content and Self-Knowledge', *Philosophical Topics* 17 (1989), 5–26; and 'Externalism and Inference', *Philosophical Issues* 2 (1992), 11–28. I think that there are multiple grounds for finding his arguments unsound. I discuss these issues briefly in

A second entitlement is involved in fulfilling the first. This is entitlement to rely on purely preservative memory. Purely preservative memory is anaphoric retention of representational content and mode type. This entitlement, too, yields immunity to brute error. One cannot be warranted in relying on purely preservative memory and make a mistake in the preservation. Memory errors in reasoning derive from allowing some other type of memory besides purely preservative memory to enter the reasoning, or from some malfunction in purely preservative memory.

A third entitlement is present in those deductive inferences that involve coming to some reason-supported conclusion. Rules for deductively valid transitions are not rules for using deductive inference to provide reasons.[83] The former rules constrain permissible reasons, but do not guarantee that premises that lead validly to a conclusion are reasons for the conclusion. The premise of a deduction could be unwarranted. Then a valid deductive transition from it would provide no reason for the conclusion.

An individual could make an inference between propositions that are in fact connected by a valid rule of inference, without relying on the logically valid connection. The individual might make the inferential connection by rote, for example. Such transitions are not really deductive inferences. They are surely not reason supporting by virtue of deduction. Such inferences are not correctly psychologically *explained* in terms of valid logical inference rules. Inferences that are correctly explained in terms of valid logical inference rules must involve, and must be further explained by reference to, exercise of a competence with the logical constants—not a rote connection, or some other non-logical connection.

For an inferential transition to provide *reason support* for a conclusion by way of a deduction, the transition must be correctly explainable in terms of deductive inference rules and competence with logical constants; the premises of the inference must be warranted; and the reasoning must be relevantly non-circular. For a deductive reason-supporting transition to be correctly explainable in terms of deductive principles, its premise must ground some rational explanation of why the conclusion is worthy of belief *for the individual*. The individual need not be able to give the explanation. But the individual's competence with logical constants must rationalize—ground a potential explanation of—why the conclusion is belief worthy for the inferrer. Reasons are answers to potential why questions. An individual's competence with logical constants must comprise

'Memory and Self-Knowledge', in Peter Ludlow and Norah Martin (eds.) *Externalism and Self-Knowledge* (Stanford: CSLI Publications, 1998). There is further literature on this topic that I will not go into here. The points in the text here supplement my 1998 discussion. For a somewhat different view of reasoning in slow-switching cases, see Mikkel Gerken, 'Conceptual Equivocation and Warrant by Reasoning', *Australasian Journal of Philosophy* 89:3 (2011), 381–400. I regard Gerken's view as broadly congenial. It could be accommodated with minor revisions of the points made here.

[83] I am using Gilbert Harman's point that rules of logic and rules of reasoning are not identical. See Harman, *Thought* (Princeton: Princeton University Press, 1973). One can engage in valid, even sound, deductive inference even if the premises provide no reasons for the conclusions.

some (non-meta-representational) competence understanding of the connection between warranted premise and inferred conclusion—a connection that is belief supporting.

Now, a point analogous to the point about entitlements to transitions as deductively valid applies to entitlements to transitions as reason transmitting or reason supporting. These entitlements again yield immunity to brute error. If a transition in an inference is made with the function of providing a reason—if the individual reasons to commitment to a conclusion that is purportedly supported by the premises, not merely deductively infers the conclusion from the premises—the individual is entitled to the transition step, even lacking an ability to think meta-representationally about it, if and only if the transition is in fact governed by (correctly explainable in terms of) correct principles of reason support.[84]

These entitlements in deductive inference and related ones in inductive inference are, I think, developmentally the first warrants that yield immunity to brute error.

I turn from this first class to a second class of warrants that yield immunity to brute error. This class is comprised of warrants to believe simple, self-evident truths on the basis of understanding them. If one understands I am now thinking, either snow is white or it is not the case that snow is white, or 2 + 2 = 4 well enough to be justified in believing it through understanding it, being warranted insures not being mistaken in the belief.

A third class of cases of immunity to brute error comprises entitlements to *believe* non-inferentially *in simple cases* that one attitude, or content together with a mode type, is a reason for another. *Simple cases* are those in which the individual uses a lower-level reasoning transition—entitled in the way of the third example in the first class—as a basis for the meta-level judgment about the transition relation. This class is the meta-representational analog of entitlements to first-order reason-supporting transitions. I distinguish belief that one thought is a good reason for another from belief about the form of the inference, and from theoretical generalizations about inference. I am concerned here purely with simple, *de re*, evaluative beliefs *about* particular transitions: *P* is good [or bad] reason for *Q* [perhaps relative to background *B*], where the individual can and does reason correctly—*p* so *q*.[85] In simple cases, if one's non-inferential understanding of the reason relation between particular attitudes is good enough to entitle one to the meta-belief that one attitude is a reason for another, the understanding is veridical: the reason-support relation holds. One can be warranted but mistaken in beliefs about reason relations. But such mistakes rest on inference in theorizing, or on complex cases. In simple cases,

[84] Rules for transitions in inductive inference are not well understood. But it is clear that if a step is warranted, it is not inductively fallacious, although it could lead to a false, warranted conclusion. Reliance on the transition rule, not the conclusion, is immune to brute error.

[85] I believe that the points apply to prima facie reasons as well as conclusive reasons.

warranted judgments are true. Errors derive from irrationality or other failures of understanding that undermine warrant.

IV

The key to understanding warrants that yield immunity to brute error does not lie in necessity. There are warrants to believe necessary truths that do not yield immunity—warrants to believe empirical necessities. There are warrants that yield immunity that do not attach to necessary truths—*cogito* cases.

The key to understanding warrants that yield immunity to brute error does not lie in apriority. Some inductively based mathematical conjectures are, I think, apriori warranted but not immune to brute error. Not all warrants that yield immunity to brute error are apriori. Entitlements to beliefs about one's sensations depend for their warranting force on having the sensation. They are not apriori. Such entitlements can still yield immunity to brute error.

What grounds epistemic warrants that yield immunity to brute error? What distinguishes these three classes of competencies? What distinguishes other competencies that yield immunity to brute error? I think that the combination of four features grounds the immunity.

First, all the relevant competencies are, or rest psychologically and epistemically on, one or another type of propositional understanding.

Second, all the relevant types of understanding, including but not limited to competence understanding, are constitutive of having certain explanatorily or normatively significant types of point of view. In order for the type of point of view to be possible at all, these types of understanding must, in their natural undamaged states, be reliable.

Third, fulfilling representational functions of the states that are, or rest on, these types of understanding either forges constitutive connections *within* the relevant type of point of view, or makes commitments that express an understanding of such constitutive connections.

Fourth, representational success in fulfilling representational functions does not depend on anything beyond exercising the understanding in a psychologically well-functioning way. In particular, being representationally successful (being veridical, preserving veridicality, and so on) on any given occasion does not depend on anything outside the psychology on that occasion, and does not depend on having any information beyond that involved in exercising, in a well-functioning way, the understanding that is constitutive to having the relevant type of point of view.

Immunity to brute error is derivable from these four features of a representational competence, together with this constitutive principle governing warrant:

> (E) A propositional state or occurrence is warranted, on a given occasion, if and only if it is the result, on that occasion, of the exercise of a representationally well-functioning propositional competence that provides a representationally

reliable, epistemically good route to veridicality, allowing for the natural limitations of the competence with respect to its subject matter, and limitations of the information available to it.[86]

By the first of the distinctive features, the relevant psychological states or occurrences rest on a reliable propositional competence. By the third and fourth, veridicality on a given occasion does not depend on extra-psychological conditions or on specific information beyond the minimal information necessary to exercise the understanding. By the second, third, and fourth features, being representationally successful (being veridical, preserving veridicality, and so on), on a given occasion, depends only on exercising the understanding of connections within the point of view—the understanding constitutive to having the relevant type of point of view—in a well-functioning way. It follows that a propositional state or occurrence is warranted if it relies only on a well-functioning exercise of a relevant type of understanding. Any malfunction of a relevant competence undermines warrant for a state or occurrence that relies on the competence. By the fourth feature, any error in a state that relies on a relevant type of understanding derives from substandard functioning of the competence—and hence undermines warrant. So any *warranted* state or occurrence that relies only on the understanding is representationally successful. In other words, reliance on any of the relevant types of understanding is immune to brute error.

Let us return to the three classes of warrant that yield immunity to brute error with this grounding in mind.

Take the first class of cases. A reliable competence understanding that consists in making valid transitions in deductive inferences that rely on purely preservative memory is constitutive to any propositional point of view, starting with the simplest empirical ones. To have states with propositional structure, one must be competent to use such structure reliably in inference, including deductive inference. Successful exercises of this competence forge constitutive connections within a propositional point of view. If the competence is exercised in a well-functioning, unhindered way, the transitions are representationally successful. They preserve truth in virtue of form. An individual is entitled to rely on well-functioning exercises of the competence—on the transitions—as deductively valid. So well-functioning exercises of the relevant competence understanding are warranted only if transitions are deductively valid.[87]

[86] For the framework that embeds (E), see my 'Perceptual Entitlement'. The point of the phrase 'allowing . . . ' in (E) is to indicate that as long as the competence is reliably veridical, and functions representationally *well* in using its available information—in a way conducive to getting things right in normal circumstances—exercise of the competence can be warranted. I assume that natural limitations include fallibility. Warranted failure with respect to veridicality derives either from natural limitations—such as natural imperfections in perceptual resolution—or from using *well* information that is in fact misleading.

[87] The rules for deductive transition that meet this description are restricted to simple cases. See the definition of '*simple cases*' above. Norms that are not explanatory of an individual's actual reasoning, because they are too complex, are not relevant to the account.

The same points apply to the competence understanding in exercising purely preservative memory. Having a naturally reliable competence of this kind is constitutive to having any point of view capable of propositional inference, indeed any representational competence at all. The competence makes constitutive connections in a representational point of view. If the competence is exercised without malfunction or hindrance, it is representationally successful in preserving mode and content from earlier states. Individuals are entitled to rely on exercises of a reliable competence that is well-functioning and unhindered. So exercises to which individuals are entitled are representationally successful in preserving mode and content. So the entitlements yield immunity to brute error.

Similarly, for competence understanding that yields reason-supporting transitions in reasoning. A reliable understanding of this sort, applied in simple cases, is constitutive to forging structural connections in any propositional point of view. Almost trivially, individuals capable of reasoning are competent to make moves between propositional steps that, relative to their background information, are in fact reason supporting—as long as their faculties do not malfunction and as long as internal conditions do not hinder natural exercise of the competence. Exercise of the understanding unhindered by disease or bias yields reason-supporting transitions. Being entitled to such transitions guarantees representational success—reason-supporting transitions. So entitlement yields immunity to brute error.

The second class of warrants that yield immunity to brute error concerns epistemic starting points, not transitions. They are otherwise similar. Understanding simple logical and mathematical truths well enough to be warranted through understanding them constitutes an explanatorily and normatively significant type of point of view. That type of point of view is a natural psychological kind. Believing the relevant truths, on the basis of understanding them, is a reliable route to truth. In fact, such beliefs are rational starting points in reasoning to other truths. Any error is a failure of the sort of understanding whose successful exercise constitutes the type of point of view. Warrant for believing simple logical and mathematical truths lies in understanding them.[88] If one is warranted in believing them, one's belief is true. So the acceptance relative to the warrant is immune to brute error.

What of the third class of cases—meta-representational understanding of reason-support relations? Competence to make reliable judgments about reasons *as* reasons is distinctive of critical reason, certainly an explanatorily and normatively significant type of point of view. Recall that *simple cases* are, by definition,

[88] Understanding self-evident logical truths is, I think, an abstraction from deductive transitions in reasoning. Understanding truths of *pure* logic is, I think, a further abstraction from instances of such principles. Understanding simple truths of arithmetic is, I think, an abstraction from counting. Understanding other simple mathematical truths, such as the simplest axioms of set theory (for example, the axiom of extensionality), is not an abstraction from particular cases but is part of a minimum explicational understanding of relevant concepts (here, the concept set). None of these truths is vacuous in the positivists' sense.

those in which the individual uses a warranted lower-level reasoning transition as a basis for the meta-level judgment about the transition relation. In simple cases, being warranted guarantees true judgments. For they function to forge certain constitutive structural connections within a point of view capable of critical reason, by recognizing constitutive structural connections within the lower-level propositional point of view. The meta-representational judgments are grounded in lower-level thinking of the reason relations *in* the meta-representational thinking about them. To get things right, the judgments need use no other information than that involved in the meta-representational understanding that conceptualizes, while relying on the competence with, the warranted lower-level reason-supporting transitions. Forging constitutive structural relations in a critically reasonable point of view just is correctly representing the warranting relations in the transitions in the lower-level reasoning that one evaluates. (As indicated in the first class of cases, the lower-level entitlements themselves yield immunity to brute error.) Any error in using the competence is a malfunction of the competence. Being warranted in using the normally reliable competence in simple cases requires that the competence function well on the occasion of use. So since any error in using the competence in simple cases is a failure in its function to establish constitutive connections within a critically reasonable point of view, any error in simple cases undermines warrant.

In understanding warrants that yield immunity to brute error, it is instructive to reflect on the contrast class. The most primitive warrant that allows brute error is entitlement to perceptual belief. Like all warrants, including those that yield immunity to brute error, perceptual entitlements are explicable only against a background of veridical states. But there are crucial differences in this background.

The first, less basic difference concerns relations between competencies underlying the warrants and having an explanatorily and epistemically significant type of point of view. An individual can have an empirical point of view, even one with warranted beliefs, and yet not be in a position to get any perceptual beliefs right.[89] How could such a thing happen?

[89] I think that something stronger applies. An individual can have perceptual beliefs, but no warranted ones. The nature of perceptual states does not require that they be reliably veridical in their normal content-determining environment. (Of course, most perceptual states *are* thus reliably veridical. See 'Perceptual Entitlement' and note 68 above.) I have been misunderstood to hold, in that very article, that it is apriori and constitutively true that perceptual systems are reliable in normal circumstances—the circumstances in which their contents were formed. I do not hold this view, and consciously wrote around it in that article. An example of the misunderstanding is Anthony Brueckner, 'Content Externalism, Entitlement, and Reasons', in Sanford C. Goldberg (ed.), *Internalism and Externalism in Semantics and Epistemology* (New York: Oxford University Press, 2007), 160–176. I had earlier been guilty of holding, without argument, that the nature of a perceptual state requires reliability. See 'Our Entitlement to Self-Knowledge', 106 n11. But I came to give up the view in the late 1990s. Obviously, if there were a good apriori argument for it, there would be a very direct and comprehensive argument against scepticism.

An individual can inherit perceptual equipment that obtained representational content through patterns of interaction, including veridical applications, in a certain environment. Call this environment '*the normal environment*' for that equipment. (See notes 68 and 89.) Suppose that when given appropriate proximal stimulation, this equipment produces types of perceptual beliefs that would be reliably veridical in the normal environment. However, any individual can be placed in a non-normal environment in which, despite similar proximal stimulation, every perceptual belief lacks the kinds of distal antecedents that gave content to the individual's perceptual equipment. Then the individual has an empirical point of view. The individual's cognitive equipment works as well as it can, and reliably, relative to the warrant-determining normal environment. Yet nearly all the individual's perceptual beliefs, including warranted ones, are false.[90]

An analogous situation is not possible with types of understanding that ground immunity to brute error. Inability to exercise those types of understanding *veridically* would make having the relevant sort of point of view impossible. An individual that could not naturally make valid or reason-supporting deductive inferences would lack a propositional point of view. An individual that lacked an understanding that, when naturally exercised, yielded belief in simple logical or mathematical truths would lack a capacity to understand logic or mathematics.[91] An individual whose natural competence did not yield true beliefs about simple reason-support relations would lack critical reason.

A second, more fundamental difference between perceptual entitlements and entitlements for types of understanding that are immune to brute error helps explain the first. The difference is that well-functioning, unhindered exercise of the perceptual competence underlying perceptual-belief formation does not insure representational success. Success on particular occasions depends on non-psychological, "brute" connections to distal matters, beyond the point of view, in the environment. The distal matters can vary so as to yield either success or failure, even while the psychological competence for forming perceptual beliefs functions well and in an unhindered way.[92]

[90] Clearly, the same sort of point can be made even if the individual is located in its normal environment. Every perceptual belief results from a proximal stimulation with artificial, non-standard distal antecedents. The deeper point is that constitutive explanation of the content of perceptual states in terms of *veridical* cases, which *is* I think apriori necessary, does not necessarily depend on the veridical cases' being the most common cases.

[91] Understanding logic and mathematics, beyond mere competence understanding, contrasts with reasoning logically or mathematically. It involves minimal explicational capacity. I think that such understanding is a stage of development that marks a significant type of point of view.

[92] See my 'Our Entitlement to Self-Knowledge'; 'Perceptual Entitlement'; 'Disjunctivism and Perceptual Psychology', *Philosophical Topics* 33:1 (Spring 2005), 1–78, especially sections VI–VII; "Disjunctivism Again', *Philosophical Explorations* 14:1 (2011), 43–80, especially sections II–IV.

Forming a given kind of perceptual belief depends only on antecedent psychological sets, psychological formation laws, and proximal stimulation.[93] Representational success depends partly, on each occasion, on distal matters—beyond the individual's psychology. Distal matters can vary while producing relevantly similar proximal stimulation. As a result, a reliable competence for forming perceptual belief that functions as well as it can in given circumstances—given the information available to it—can produce warranted but false belief—brute error. No analogous distal, subject-matter variation can occur with respect to well-functioning warranted exercises of the sorts of understanding that are immune to brute error. With such understanding, the constitutive connections—and the connections that representational success depends on—are entirely within the relevant type of point of view. Moreover, exercising the well-functioning understanding that is constitutive of a relevant type of point of view just *is* successfully traversing, preserving, or recognizing the relevant connections within the point of view. Getting things right does not depend on conditions outside the psychology and does not require information beyond that which is necessary to exercise the understanding. Well-functioning, unhindered, warranted reliance on the competence yields representational success on each particular occasion and in any environment. So well-functioning, warranted exercises are immune to brute error.[94]

Of course, default entitlements to perceptual belief are not the only entitlements that allow brute error. Similar points apply, however, to entitlements to empirical substantive memory and to anticipation of one's bodily action or external events. Mistakes can be warranted because they depend on non-normal relations to the environment on given occasions. One can exercise the reliable competence unhindered and be mistaken.

An interestingly different class of warrants that allow brute error comprises the products of various types of inductive reasoning. I have three types of induction in mind: empirical inductive reasoning about the environment, for example, inference to the best explanation from perceptual beliefs; inductive reasoning to conjectures in mathematics; and inductive reasoning about oneself. The conclusions of these types of reasoning can be warranted (justified), but mistaken. For example, in forming well-reasoned judgments about one's

[93] This is a statement of the Proximality Principle, which guides the science of perceptual psychology. See my 'Disjunctivism and Perceptual Psychology', section IV; *Origins of Objectivity*, 385ff.; 'Disjunctivism Again', sections II–III, V. Other issues in this paragraph are discussed in "Perceptual Entitlement."

[94] The fact that a reliable competence represents only internal psychological elements does not suffice to make the representational states that rely on it immune to brute error, relative to warrants for them. Reliable inductive reasoning from warranted premises about one's psychological states can yield warranted but mistaken beliefs. Representational competencies that are basic constitutive elements or basic constitutive connections in a point of view are immune to brute error. Such competencies include purely preservative memory, inferential transitions, using reasons, understanding basic logical and mathematical truths, recognizing simple reason-support relations as such, and the sorts of self-understanding that I will discuss.

character, one can be mistaken. The induction might be based on behavioral evidence, or on types of self-knowledge that *can* be arrived at through Self-Understanding.

Inductive reasoning uses patterns found in a base case, or cases, to form judgments about a further case, or cases—relying on the likelihood that the pattern reapplies. Well-functioning, warranted use of induction never insures veridicality of the conclusion.[95] It is inherent in the method that even if the beliefs that form the induction base are true, the conclusion can be justified but mistaken.

In all these cases of capacities that are vulnerable to brute error, representational success—veridicality—depends, on each occasion, on more than optimal representational functioning of psychological competence, given the information necessary to exercise it at all. Not so, with the competencies that yield immunity to brute error. Although relying on them can yield failure of veridicality, such fault lies in themselves, not in a brute failure to match a subject matter beyond them.

V

By reflecting on other cases of warrants that yield immunity to brute error, I hoped to illuminate Self-Understanding. I now return to that topic. Self-Understanding is constitutive to having an apperceptive rationally accessible point of view. Immunity to brute error in Self-Understanding follows the pattern established by other examples of immunity. To explicate entitlement to Self-Understanding and Self-Knowledge in detail, I must elaborate structural considerations set out at the end of Lecture II.

Self-Understanding and Self-Knowledge involve predication of a psychological element to oneself. I distinguish several aspects of the predication. These elements mark specific competencies that ground specific warrants.

Consider the predication in I believe that Mozart loved Haydn. I refer to myself with a singular application of the self-concept I, and I attributively apply the concept believe that Mozart loved Haydn to myself. This concept consists of three primary components—the verb-concept believe that indicates a psychological mode (belief), the present tense of the verb-concept, and the canonical singular concept that Mozart loved Haydn that denotes a representational content. I shall say that the full concept, believe that Mozart loved Haydn, *indicates* a *type* or *kind* of psychological element—belief that Mozart loved Haydn. It also thereby indicates the mode type, belief.

In veridical predications in Self-Understanding, the attribution establishes a *de re* relation to *instances* of those types—here, an instance of a belief that

[95] I exclude mathematical induction that occurs in Peano arithmetic. I think that warrant that derives from relying on such induction does yield immunity to brute error.

Mozart loved Haydn and an instance of the mode, belief. I shall say that, in such cases, the predicate concept that indicates the psychological type or kind *betokens* the instance of the type or kind. So the predicate concept *indicates* the psychological kind and *betokens* the instance of the kind. Both indication and betokening are types of representation.[96]

Predication in Self-Understanding involves attribution as well as indication and betokening. When attribution is veridical, the indicated *kind* is attributed, and the betokened *instance* is also attributed. These are different attribution relations, but I will call them both '*attribution*'. In Self-Understanding, kind and instance are attributed to the individual picked out by application of the singular self-concept. Although all these elements work together in the relevant predications, they are different aspects of the predication and correspond to different aspects of understanding. Thus there is *indication understanding*, *betokening understanding*, and *attribution understanding*.

Betokening is necessarily guided by indications of an attribute—in veridical Self-Understanding, the indicated kind. Indication understanding is always a component of betokening understanding. I will not discuss indication understanding separately.

Betokening understanding is usefully thought of as a *vertical* relation. It is a relation between a higher-level understanding and a lower-level psychological instance that is understood. Both understanding and what is understood are elements in a point of view. The understanding establishes the sort of multi-tier point of view discussed earlier.

My main task is to explain wherein being relevantly warranted in betokening understanding entails *correct* betokening understanding. The explanation should account for the warrant's intuitively yielding immunity to brute error. I begin with Self-Understanding of *propositional* states and occurrences. Later I consider other types of Self-Understanding.

In tracking what follows, one must bear firmly in mind a distinction between lower-level competence understanding and meta-representational understanding. *Lower-level competence understanding* is just a competence to think with propositional contents. *Meta-representational understanding* is a competence to think about the exercise of the lower-level competence, or the modes and contents that are instantiated in the lower-level competence.

In the meta-representational betokening understanding integral to Self-Understanding, a propositional state *instance* or *occurrence* is betokened *de re*. The instance is betokened via its constitutive aspects—its representational content and its mode. In attributing each of these aspects, meta-representational betokening understanding makes essential use of lower-level competence understanding.

[96] Betokening is the predicative analog of syntactically singular *de re* reference. Not all predications involve betokening. The predication 'Some people love others' does not, if the quantifier is not associated with *de re* thoughts about particular instances of loving.

Recall from the previous lecture that in attributing representational content to one's propositional state- and occurrence-instances, one must exercise lower-level competence understanding in the represented thought content as one uses the canonical representation that names that content.[97] An under-recognized aspect of betokening understanding is that it involves an analogous phenomenon in use of the predicate for the *attitude mode* of the propositional content. Application of the concept believes in betokening understanding does not use a *represented* representational content. For believes does not indicate a representational content. But one can preserve in memory, with application of believes, the belief mode; and one can retain dispositions to implement reasons connected to any given belief. That is, in betokening understanding of an attitude instance, one exercises lower-level memory that preserves representational content—of which one has lower-level competence understanding—and preserves sensitivity to its being believed.[98] Further, in understanding the belief, one can prime dispositions to associate it with other attitudes that bear lower-level reason connections to it. In betokening understanding, one uses the lower-level preservation through memory of the attitude mode, along with the content, in picking it out, *de re*. One retains in memory the belief mode as one represents it. I will reflect, first, on aspects of this meta-representational competence and, then, on epistemic warrants for its uses.

Purely preservative memory functions to preserve mode type and representational content, without adding warrant or new subject matter to current psychological transactions. (See note 63.) Fulfilling this function by purely preservative memory is constitutive to *any* representational psychology—propositional or not, meta-representational or not. For example, in propositional inference the mode and content of earlier steps must be preserved for later use in inference. Or in a use of visual perception in action, purely preservative memory retains the mode and representational content of the visual state. Although one is entitled to rely on purely preservative memory, and although such reliance is necessary for the success of the inference or perception-based action, such reliance is not part of the justification or a practical warrant for the action. Warranted reliance on purely preservative memory is a necessary enabling condition for the conclusion's, or action's, being warranted. My account of the epistemic warrants for Self-Understanding and Self-Knowledge in non-*cogito* cases hinges on reflecting on preservational capacities, central to any representational psychology, and

[97] For more on the canonical representation of representational content, see my 'Individualism and Self-Knowledge'; 'Postscript: Frege and the Hierarchy', in *Truth, Thought, Reason: Essays on Frege* (New York: Oxford University Press, 2005); 'Postscript: "Belief *De Re*"', in *Foundations of Mind*; and 'Five Theses on *De Re* States and Attitudes', section v.

[98] I write 'preserves sensitivity' because the lower-level purely preservative memory does not preserve any representation of a belief as such. Lower-level aspects of a psychology are not meta-psychological. They nevertheless track differences in the modes of past steps in inferences. Beliefs function differently in the psychology from suppositions. Purely preservative memory tracks sensitivities to different modes, and in that sense preserves the mode types.

their role in meta-representational understanding—especially betokening understanding.

There are importantly different types of Self-Understanding and Self-Knowledge. Preservation, though not purely preservative memory, figures in all of them. Preservation is necessary to Self-Understanding from the inside, and to the role of Self-Understanding in providing buck-stopping status to psychological elements in critical reason and morality. It is part of the constitutive tissue of selfhood. So understanding the forms that preservation takes between lower-level and meta-psychological elements lies at the heart of my project.

In all cases, preservation between the lower-level understanding and the meta-representational understanding *of the representational content* works in the same way. In propositional cases, the lower-level thought content is always preserved in the meta-representational thought by being *used*—thought—*in* the canonical name for the content. I will concentrate on various types of preservation of *mode type* and *mode instance*.

Preservation between lower-level thought and meta-representation of it occurs trivially in pure *cogito* cases. No memory is involved. In pure—formally self-verifying—*cogito* cases (I [hereby] entertain the thought that writing requires concentration), the mode—entertaining—of the lower-level thought (writing requires concentration) is inevitably preserved in any thinking of the meta-representational thought, no matter what its mode is. Entertaining the lower-level thought content is guaranteed in the meta-representational thought. Even the instance of entertaining is referred to reflexively in the thinking of *cogito*. Pure *cogito* cases are infallible. If one thinks a thought of that form, one thinks a truth.

In impure *cogito* cases (I [hereby] judge that writing requires concentration), the indicated lower-level mode (judgment) is not entailed to be the mode of the content cited by the underlining. In such a *cogito* thought, one could judge that one is judging the lower-level content, and be mistaken about one's making a judgment. The two levels of judgment could come apart. One could judge that one is judging that writing requires concentration, although one does not thereby *judge* that writing requires concentration. The latter content (that writing requires concentration) might pass non-committally through one's mind. Such cases are pathological misuses of the *cogito* form of thought. Normally, the meta-representational judgment coordinates with a performative judgment of the lower-level content. Normally, the lower-level judgment is a constituent of the meta-representational act of judgment. Then the lower-level mode and its instance are preserved in the meta-representational judgment not by entailment, but by coordinated, performative, reflexive agency.[99]

[99] For more on *cogito* cases, see my 'Individualism and Self-Knowledge'; and 'Mental Agency and Authoritative Self-Knowledge: Reply to Kobes', in Martin Hahn and Bjørn Ramberg (eds.), *Reflections and Replies: Essays on the Philosophy of Tyler Burge* (Cambridge, Mass.: MIT, 2003),

Cogito judgments concern present occurrences. Preserving mode type and mode instance is trivial in pure cases, and almost trivial in impure cases. It does not depend on a causal connection between the understood psychological element and the understanding. The connection is contemporaneous. Preservation of mode in judgments that realize Self-Understanding of one's present, continuing, *standing* attitudinal *states* (belief, intention) and certain among one's past psychological states and episodes is more interesting.

How does preservation work in Self-Understanding of standing belief? Memory often figures in a present-tense judgment like I believe that Mozart loved Haydn. Purely preservative memory retains a standing belief—standing from before the judgment. The judgment takes the belief not to have just formed. Purely preservative memory preserves content and mode of the belief for later use. In a meta-representational psychology, the lower-level, purely preservative memory is supplemented by conceptualized representation of both the preserved representational content and the preserved lower-level mode—here, belief. In conceptualizing these matters and *representing* the belief as the individual's own, the individual indicates not only the mode type *belief*; the individual also thinks *de re* of the mode instance. The individual represents his or her own instance of that type of belief in the judgment. When combined with canonical understanding of representational content, such powers yield betokening understanding of instances of standing propositional states.

The memory used in this sort of recognition is meta-psychological autobiographical memory. But it is not episodic memory. The memory is purely preservative memory, supplemented with a meta-representational capacity (a) to conceptualize the tracked preserved mode—type and instance—and representational content of the standing state, and (b) to recognize that it is a state that has been standing from the past.

A similar account applies to Self-Understanding of some past psychological events, especially relatively recent ones. Meta-psychological autobiographical episodic memory again relies on purely preservative memory, supplementing it by conceptualizing the representational content and mode (type and instance) of the past psychological event.[100]

Some Self-Understanding and Self-Knowledge are forward looking. One can be entitled to believe, fallibly but non-empirically, that one will retain a present intention or belief some moments, or longer, into the future. One can anticipate that a specific standing state type and state instance will remain standing into at least the near future, if there is no reason to doubt. Forward tracks of preservation

417–434. In the latter article, I discuss impure performative *cogito* judgments involving modes like intention as well as modes like judgment. See note 74.

[100] Preservation of mode in memory need not maintain commitment involved in the mode. One might have a meta-psychological episodic memory of one's having judged that so and so, even as one currently doubts so and so. Past attitude instances are preserved, inasmuch as betokening understanding relies on lower-level preservational memory to track that commitment.

are as basic to representational psychologies as backward, causally sustained, memory tracks. When utilized by meta-representational powers, including canonical representation of representational content, they figure in *de re* understanding of the mode type and mode instance of future representational states.

Use of these types of preservation—of a present event, of a standing state or a past episode through purely preservative memory, and of a future anticipated state or event—is constitutive to being a self. An individual that lacked any one of these preservational capacities would lack powers integral to applying critical reason in inference or action. I made an analogous point in Lecture II regarding the role of meta-psychological autobiographical episodic memory in critical inference, decision making, and dialectic.

VI

I turn now to *epistemic warrant* for betokening understanding in Self-Understanding. Epistemic norms are grounded in competencies to realize representational functions. Realizing representational functions involves forming and preserving veridical psychological states and occurrences. A propositional attitude is epistemically warranted only if it is the product of a competence that realizes its representational function reliably and well, given the individual's perspective. Powers to produce and preserve veridical representation—representation of mode types and mode instances, and of representational contents—must, in their natural state, be reliably veridical if they are to be constitutive aspects of selves and help ground critical reason. So reliance on these powers by selves—critical reasoners—is epistemically warranted. In particular, in purporting to single out *de re* one's propositional attitude instances, one is default warranted in betokening understanding, if one bases judgments about the instances on the relevant types of preservation.

The judgments based on the preservational routes just discussed are immune to brute error. The meta-representational understanding relies on lower-level preservational capacities in specifying representational content, mode type, and mode instance. The warrants that yield immunity to brute error are all based on this reliance of meta-representational understanding on the lower-level competence understanding, perhaps supplemented by purely preservative memory or anticipation. I will apply these epistemic points to each of the types of Self-Understanding that I have discussed.

Pure *cogito* thinkings are warranted through self-evidence. Individuals are warranted by understanding their thinkings. Such warrants yield immunity to *any* error. Warrant for impure *cogito* judgments lies in *de re* understanding. Understanding coordinates the lower-level performative act with the higher-level *cogito* judgment. Impure *cogito* judgments are not immune to error. Coordination failures that yield error are pathological and undermine warrant. If one judges

I am hereby judging that Mozart loved Haydn and fails to judge that Mozart loved Haydn, one's self-understanding is pathological; warrant lapses. One misuses such judgments if one does not engage in the act that one judges oneself to be engaging in. Such misuses undermine warrant. Impure *cogito* cases are immune to brute error.

Errors in judgments, based on Self-Understanding, about standing propositional attitudes, past or present, derive from malfunction of purely preservative memory or from relying on other capacities besides purely preservative memory. Then the judgment either is unwarranted or is backed by a different warrant from the one that backs Self-Understanding.

Suppose that one applies the concept intend to finish the project or the concept believe that most domesticated dogs are friendly in a purported betokening understanding. Suppose that one lacks the attributed psychological state. The first-level purely preservative memory could have malfunctioned. Or it could have been derailed by competing psychological needs, or other sources of self-deception. Malfunction or relying—however unconsciously—on factors that are irrelevant to supporting a belief undermine the relevant warrant (entitlement). Or one could form a warranted but mistaken judgment with the same content, inferring it from behavioral evidence. Then one's warrant would be different. In all relevant cases of error, either one's default entitlement for the judgment about one's standing propositional attitude lapses, or the judgment does not rest on the *relevant* entitlement. So there is no brute error in judgments warranted in the relevant way. The key route is a conceptualization of mode and content that is preserved by purely preservative memory. The capacity to track one's standing states by way of their mode and content, and to represent them as having been held (not presently formed), is a basic preservational power in a critical reasoner. Minimal exercise of this power yields immunity to brute error.

A similar line applies, I think, to some judgments about past psychological events—judgments based on meta-psychological autobiographical episodic memory that relies on lower-level purely preservative memory. Suppose that one judges I was just then thinking that writing requires concentration. The judgment rests on meta-psychological autobiographical episodic memory. The memory conceptualizes a mode and content preserved by the basic transtemporal psychological connector—purely preservative memory.

I claim that relying on betokening understanding that utilizes purely preservative memory in Self-Understanding of both some standing propositional attitude states and some past propositional attitude events is immune to brute error. I want to elaborate this claim and defend it. First, again, I discuss the warrant for relying on purely preservative memory in these types of Self-Understanding.

Purely preservative memory functions just to retain the mode and representational content of psychological states. It does not represent anything. It only connects. It preserves representation from other states. Relying on it is warranted non-empirically, though, of course, it can preserve empirical content. It is a

constitutive precondition of any inference. Its being naturally reliable is a condition on having a functioning representational psychology.

Purely preservative memory is fallible. It can lapse momentarily, or become unreliable, through age or disease. Short-term forms do not operate after their decay times. Longer-term forms, which, as far as is known, lack built-in time limits, not only fail on occasions but deteriorate in reliability. Still, whenever purely preservative memory fails to retain mode or content, the failure is a malfunction. Errors are not attributable to a representationally well-functioning preservative competence. Any malfunction undermines warrant.

Of course, error can derive from other psychological factors' over-riding purely preservative memory, or being used instead of it. Then errors do not derive from purely preservative memory. Since the individual relies on other factors, warranted or not, any such errors do not occur despite warranted reliance on purely preservative memory.

How and whether purely preservative memory interacts with other psychological factors, in general or on any given occasion, is an empirical matter. But when it does not malfunction and is not over-ridden by other psychological factors, it preserves mode and content of past representational states. Where there are errors that are attributable to *its* operations, the errors are not brute errors. They are errors of malfunction that undermine warrant. That is why reliance on purely preservative memory is immune to brute error.

Are other types of memory, besides purely preservative memory, subject to brute error? In a sense, perceptual memory inherits perception's—and perceptual belief 's—vulnerability to brute error. A given memory could be brute mistaken because it derives from a perception that is brute mistaken. However, here, the vulnerability is the vulnerability of *perception* to brute error. Errors specifically attributable to memory can seem to result from memory's not making the best cognitive contribution that it can. And it may seem that when perceptual memory undergoes such failure, warrant for relying on it lapses, because memory would on such occasions not operate well, even if it is generally reliable. Such failures would differentiate memory from perception. Perception can operate optimally given its proximal input; but because of abnormal distal conditions, it can still go wrong. Cases in which one is warranted in relying on it are cases in which it operates well, given the information that it has. If inaccuracy is correctly attributable to the operation of the perceptual system in a given case, warrant for relying on perception lapses. It can seem that where error is attributable specifically to perceptual memory, warrant lapses. So it can seem unclear how brute error in *any* memory is possible.

This account is too simple. It is known that some memory—especially propositional, longer-term episodic memory—involves constructing narratives and carrying out implicit (unconscious) quasi-inductions about the past. These constructions are partly based on other things that we believe. They are fairly reliable in providing important general attributes of remembered episodes. But with

respect to matters in the relatively distant past, memory is often strikingly unreliable in filling in specific details.[101]

One can make sense of brute error attributable to such memory. To be warranted, the memory type must be reliable with respect to the level of detail remembered in the case in which the memory is employed. One would not be entitled to rely on purported memory of certain types of details that are purportedly remembered in the distant past, if memory in such cases is not reliable. But in cases in which memory is reliable in constructing the general attributes of an episode, it could fall into brute error despite following warranted operations from warranted starting points.

I make three points regarding this conjecture about brute error in episodic memory. First, to count as memory of an episode, a constructionist memory must use not only general background beliefs, but non-constructed, non-quasi-inductively produced representations of the episode. (The non-constructed memories are the singular bases for the quasi-induction.) I doubt that these singular starting points for the memory construction could be subject to brute error. Either they are well formed and accurate, or they are inaccurate by virtue of a non-constructed retention failure by the generally reliable competence. Such retention failures are just the sort of failures in representational function that undermine warrant.

Second, if there were no memories—whether or not these are singular bases for constructionist memories—of one's near-term propositional states and events that were immune to brute error, one's reflexive reviews of such states and events would never give the reviewed states and events buck-stopping status in critical reasoning and moral reasoning—by the arguments I gave in section IV of Lecture II and section I of this lecture. We have good reflective reason to believe that reflective review does give reviewed states buck-stopping status in uses of inference, practical decision making, and dialectic.

Third, the idea that near-term episodic memories always depend on matters beyond the well-functioning of one's memory—and are always subject to brute error—is implausible on its face. The idea that I was just thinking then [five seconds ago] that writing requires concentration or I am carrying out my decision [just made] to help him is always subject to error that could result from memory that functions well in the given circumstances is implausible on its face. Assume that the thought is based on conscious memory of the event, and not some collateral conscious thoughts about the past. To be subject to brute error the memory-based thought must be warranted. So the episodic memory must be generally reliable in that type of case. Error must not derive from contextual failure in the well-functioning operation of memory. For such failure would undermine warrant. Holding that memory of near-term past thinkings is always

[101] See Daniel Schachter, *Searching for Memory: The Brain, the Mind, and the Past* (New York: Basic Books, 1996). Schachter notes that John Dean's, presumably honest, detailed accounts of particular episodes were shown by the Nixon tapes to be massively inaccurate regarding details of the episodes, but accurate on their most important general features.

subject to errors that cast no aspersions on the representational well-functioning of memory is extremely implausible.

I think that our understanding of diachronic reasoning in inference, practical decision making, and dialectic indicates that in such reasoning episodic memory can utilize purely preservative memory—a type of what is called 'semantic memory' in the psychological literature. Such memory is the non-constructionist, anaphoric type of retention required in any diachronic representational operation—including inductive inference and narrative. Our understanding of the epistemic norms governing memory-based meta-psychological thoughts about one's very recent thinkings indicates that errors in such thoughts are attributable to representational malfunctions of memory. Such malfunctions, on given occasions, undermine warrant. Since I think that a natural competence to retain thought content and occurrent thought events over at least short times is constitutive to critical reason in temporal beings, I think that in us, and in any critical reasoner, there are cases of near-term meta-psychological autobiographical episodic memory in which brute error is not possible. I think that any individual that is subject to brute error in all memories of near-term thinkings would lack the cognitive equipment that is constitutive to the use of critical reason in inference, practical decision making, and dialectical thinking.[102]

I have been discussing warrants for relying on purely preservative and episodic memory in Self-Understanding of standing propositional states and past propositional events. Those warrants are aspects of the warrant for betokening understanding in these cases. I want now to focus more specifically on the warrant for betokening understanding in these types of Self-Understanding. In these cases, betokening understanding relies on purely preservative memory in Self-Understanding to retain mode and content of standing or past psychological elements.[103] Betokening understanding supplements this reliance with use of meta-representational competencies to *represent* the preserved representational content in a canonical way and to *represent* the preserved mode instance.

[102] If it is possible that there be beings that carry out critical reason in time periods shorter than those allowed for in current views about long-term memory—of which episodic memory is a subtype—then I think that a competence for an analog of meta-psychological autobiographical episodic memory would have to occur in those beings. The analog would differ only in being short-term memory. It would be memory of reasoning episodes in the very *very* recent—technically, short-term—past. Those analogs would have to be immune to brute error. Any being with representational competencies must, I think, have at least short-term memory. And beings with critical reason must be able to utilize autobiographical meta-psychological memory that retains not only content in reasoning, but retains *de re* memory of episodes of reasoning, individuated by their mode and representational content. Whether the memory is technically long term is, I think, not a constitutive matter. But in all actual critical reasoners that we know of, the relevant memory is surely sometimes long term—a few seconds or longer.

[103] In Self-Understanding of past psychological events, purely preservative memory joins, as we have seen, with episodic memory. In what follows, I focus on the relation between betokening understanding and purely preservative memory. It should be understood that what I say is meant to carry over to the cases of Self-Understanding in which episodic memory is involved as well.

Purely preservative memory preserves mode type and representational content, but *represents* neither as a subject matter. *A fortiori* it does not represent the events or state instances whose mode and content it preserves. Still, it always preserves the mode type and representational content *of a particular antecedent state instance or event*. In Self-Understanding, betokening understanding capitalizes on this fact to represent not only the mode type and representational content, but also the psychological state instance or event—which is always preserved, "there" to be referred to, when purely preservative memory functions well. Betokening understanding *uses* the lower-level purely preservative memory while *representing* what it preserves. This meta-representation is *de re* and is immune to brute error. So, when betokening understanding relies on just these competencies, either it is warranted and veridical, or it goes wrong through malfunction and warrant is undermined. Thus the basic warrant for relying on betokening understanding is immune to brute error.

What are the epistemic bases for my claim that betokening understanding in Self-Understanding of *some* propositional attitude states and events is immune to brute error?

I noted that we know non-empirically, by understanding their reflexive structure, that pure and impure *cogito* judgments involve no inference. Pure *cogito* cases are infallible. By reflecting on the form and content of the judgments, one can recognize that in impure *cogito* cases only a pathological failure to hold meta-representational and lower-level attitude modes together in a judgment could yield error, if one thinks the thoughts in a reflexive, "hereby" way. Errors derive from odd misuses of a form that one can use successfully. Although these are easy cases, they provide a model for understanding a wider range of cases.

What is the epistemic basis for my meta-meta-claim that betokening understanding in Self-Understanding of *some* past propositional events, and *some* standing attitude instances, as standing, is immune to brute error? The betokening understanding consists in a meta-representational conceptualization of the products of purely preservative memory.

The point that relying on *purely preservative memory* is immune to brute error seems secure for reasons already given. The point depends on reflecting on the function of such memory in any representational psychology. A reliable capacity to preserve mode and content is constitutive to having a system that could count as a representational psychology. Purely preservative memory is certainly not a matter of induction. Like deduction, induction *presupposes* a prior non-inferential capacity for purely preservative memory. The retention does not involve representation of what is retained. Error in retention must consist in representational malfunction of the competence, not brute mismatch with a subject matter.

What of the epistemic status of meta-psychological conceptualizations of representational content and mode (type and instance) of the products of purely preservative memory? Canonical meta-representational conceptualizations of *representational content* seem to be exercises of a competence understanding of the content together with competence understanding of the that-clause-like

canonical names. Those are competencies that we clearly have. Reflection on them shows at least as clearly as reflection on *cogito* competencies does that they are non-empirical and immune to brute error. The key issue concerns the status of my claim regarding conceptualization of *mode—type* and *instance*—of a state with a specific representational content in meta-psychological memory. How can we know that such conceptualization is immune to brute error?

In Lecture II, section IV, I discussed three ways in which meta-psychological autobiographical episodic memory from the inside figures constitutively in critical reasoning: in inference, in practical decision making, and in dialectic. In such exercises of critical reason, we can recognize conceptualization of a mode—type and instance—of a propositional attitude that is preserved for a rational project (inference, decision, dialectic).

The following principle is a norm of critical reasoning in these enterprises:

> (CR) If, in critical reasoning, one correctly and with warrant judges that a lower-level state is (or is not) reasonable, then it rationally follows directly that one has reason to sustain (or change) the lower-level state.[104]

(CR) is like an inference rule, except that it explicitly concerns transmission of reasons, not preservation of truth. (CR) implies that in the relevant uses of critical reasoning, one's Self-Understanding of the lower-level state instance is immune to brute error. For if an application were subject to brute error, a reason that is attributed, correctly and with warrant, at the meta-psychological level of critical reasoning would not transmit immediately to the lower-level state. The transmission would *always* be subject to the proviso that one's completely warranted meta-level evaluation and self-attribution is not brute mistaken. If brute error were always possible in one's self-understanding, the lower-level attitude would not have buck-stopping status by virtue of being reviewed or reviewable. In fact, *no* attitude, except in *cogito* cases, would have buck-stopping status by virtue of being reviewed or reviewable. In critical reason, the fundamental, buck-stopping states include both the meta-level review and the critically reviewed lower-level states. The buck-stopping status of the lower-level states in critical reasoning partly depends on the potential for meta-level review. They are not separable. If (CR) were not true and applicable in the three cases, critical reason would never occur outside the immediate present of *cogito* cases. Thus I think that the view that understanding our standing propositional attitudes, as having stood, and understanding of some—at least near-term—past propositional occurrences is *always* subject to brute error amounts to a scepticism about our being diachronically extended critical reasoners.

I think that we can know that (CR) governs our reasoning in the way that we know that our reasoning is governed by specific rules of deductive inference—or by specific principles governing simple deductive reason-support relations.

[104] See my 'Our Entitlement to Self-Knowledge', 109ff; 81ff in this volume.

Knowing the principle is based on reflecting on the practice and norms of critical reasoning.

I think that we can know, by reflection, (CR)'s application to some actual particular cases of critical reasoning. That is, I think that in some cases, we can know by reflection that, from a warranted meta-level judgment that a given attitude is reasonable or unreasonable, we immediately have a reason to maintain or change the lower-level attitude. We have a reliable and warranted capacity to recognize the form of transmission of reasons as they occur in some diachronic cases. This capacity is analogous to recognizing in a deductive inference, on a particular occasion, that if a premise is warranted, it provides a reason for the conclusion.

Our ability to recognize immediate reason transmission in given cases is fallible and defeasible. But it seems to me that, as in the case of knowing (CR) itself, the positive warrant for knowing some of its applications is non-empirical. The warrant is for applications of a meta-meta-representational capacity, a capacity to understand our tracking representational content and mode (type and instance) of our diachronic conscious reasoning.

It is true that the unconscious mechanisms that underlie memory in these cases are open only to empirical investigation. But the conscious states are not identical with unconscious counterparts, even those that share *representata*. For the modes of presentation and hence identities of conscious states are different. Consciousness is an aspect of the mode of presentation, hence identity, of the states and processes that I am centering on. I think that the idea that we can know non-empirically the form of reason transmission in particular cases is no less plausible than the idea that we can know non-empirically the form of deductive inferences in particular cases, or that we can know of one type of thought that it is, if warranted, a reason for another. Such knowledge is defeasible.[105] But it is apriori. The idea that we have it seems to me very powerfully plausible. Such knowledge of norms places a prima facie constraint on empirical psychological investigations.

Not all memory in self-knowledge about past psychological states is immune to brute error. Most memories with warrants immune to brute error are nearer term, even if they are technically long-term memories. As time spans lengthen, use of induction about the past and memory that makes essential use of cobbling together different experiential memories become more prevalent. Longer-term memories often rely on factors other than purely preservative memory to access past psychological events—factors subject to brute error. My thesis is that we, as selves and critical reasoners, do sometimes rely on meta-psychological

[105] The claim that we have such apriori knowledge is defeasible. Most apriori warrants for Self-Knowledge are themselves defeasible. Hume was sceptical of reasons, as opposed to natural associations. He tended to assimilate all reasoning to induction, which he construed as blind habitual association. I join the later history of philosophy in rejecting these doubts. I think that some meta-knowledge that we engage in deductive inference is probably not defeasible.

autobiographical episodic memory, supported by purely preservative memory, in calling up past thoughts. I leave open in what cases we do so.

In sum, a betokening understanding that connects *de re* applications to preserved mode instances, and to their representational contents, relies on a lower-level capacity, purely preservative memory, that is constitutive to reason. The betokening understanding itself is a constitutive connector in a psychology capable of *critical* reason.[106] An individual is entitled to such betokening understanding, and to judgments based on it, because (a) he or she is entitled to rely on the lower-level capacity, and (b) he or she is entitled to *de re* meta-conceptualization of the mode and content of attitude instances used in inference, decision making, and dialectic in critical reasoning. Moreover, we have a meta-meta-representational ability to recognize applications of the capacity cited in (b). Judgments that rely on this betokening understanding are immune to brute error, because any error in such judgments derives from failure of purely preservative memory, failure of a *de re* power of understanding constitutive to critical reason, or failure to base the judgment on those capacities.[107]

Analogous points apply to anticipatory betokening understanding. Such understanding of future psychological instances is meta-representational conceptualization that piggybacks on lower-level preservational capacities for control and continuity in states like intentions—for example, in decision making. The default entitlements attach to an understanding that is a basic connector in critical reason. They yield immunity to brute error.

Let me survey whence we have come. I have focused on betokening understanding of propositional states and occurrences. Such understanding is integral to an apperceptive rationally accessible point of view, hence to being a self. Apperception combines betokening understanding with a self's self-attribution of the understood psychological elements. It comprises the multi-tier perspective of critical reason that marks selves. Betokening understanding is also essential to beings that are subject to moral norms. It is a meta-representational capacity that, except in *cogito*-like cases, uses lower-level transtemporal preservational powers to guide *de re* representation of propositional state instances and propositional occurrences. Natural reliability of the lower-level preservational capacities is constitutive to any representational psychology, hence to reason. Natural reliability of the meta-representational betokening understanding that uses those lower-level capacities preserves Self-Understood state instances and events in a way that is constitutive to selves and to psychologies capable of critical reason and morality. Such preservation is basic to understanding one's states from the inside and to the buck-stopping status of Self-Understandable states in critical reason and morality. Epistemic warrants attach to betokening understanding because it is reliable. Successful representation, on particular occasions,

[106] See discussion of meta-psychological autobiographical episodic memory in Lecture II.

[107] I re-emphasize that this failure could involve relying on inference or on considerations that are relevant to supplementing or overriding the basic entitlement.

depends only on its good use. It does not depend on matching a subject matter outside the psychology, or on information beyond the minimum necessary to exercise the understanding. Being warranted follows from using a well-functioning, reliable, epistemically viable competence, relative to information available to it. (See (E) above.) Such uses are certainly epistemically viable and, within their appropriate domain, reliable. And the minimum information needed to use betokening understanding suffices to provide veridical memory. So when the understanding does not malfunction and is not misused, it is warranted. Error derives only from malfunction, from misuse, or from relying on other capacities. So betokening understanding is immune to brute error, relative to this warrant.

Fulfilling epistemic norms governing Self-Knowledge thus entails exercising powers constitutive to *critical reasoning*. They are not norms that are merely necessary to critical reasoning, but constitutively independent of it. Fulfilling them requires using preservational powers that are partly constitutive of critical reason. For betokening understanding of attitudes constitutively uses meta-representation guided by preservational powers that are constitutive to any reasoning.[108] Conversely, such understanding is constitutive to the multi-tier,

[108] These points and the next text paragraphs address an interesting objection by Peacocke, 'Our Entitlement to Self-Knowledge: Entitlement, Self-Knowledge, and Conceptual Re-Deployment', *Proceedings of the Aristotelian Society* 96 (1996), 117–158. Peacocke objected to my account in 'Our Entitlement to Self-Knowledge'. He held that I had successfully shown only that critical reason requires Self-Knowledge, not that critical reason figures constitutively in the Self-Knowledge. He thought that Self-Knowledge is constitutively independent of norms of critical reason, and that mastery of the concept of belief and other attitudes, mastery of concepts of representational contents, and sensitivity to first-order norms of reason "generate" an explanation of Self-Knowledge. My 1996 paper did not focus on this issue; and Peacocke's objection had some resonance, although I believe that my account of the structure of critical reason even then suggested why it was plausible to hold that Self-Knowledge in critical reasoners uses essential preservative routes that are distinctive to critical reasoning. I have maintained that rational capacities that figure in critical reasoning are partly constitutive of Self-Knowledge.

Of course, I do not hold that Self-Knowledge is *based on* critical reasoning. Except in *cogito* cases, it is not based on reasons at all. My view is rather, first, that preservative routes essential to any reasoning, hence critical reasoning, are essential to Self-Knowledge. So Self-Knowledge cannot be constitutively independent of norms of critical reason. Perhaps this point is implicit in Peacocke's own appeal to sensitivity to first-order norms of reason. Second, I think that competence to conceive reasons as such is partly constitutive of mastering concepts of belief and other attitudes. See the discussion, in section VII below, of using reason-support relations as a supplementary factor in betokening understanding. Both points bring out ways in which horizontal and vertical reason-preservation and reason-transmission routes that are constitutive of critical reasoning also figure constitutively in Self-Knowledge. For further discussion, which however does not advance as far as the present account, see my 'A Century of Deflation and a Moment about Self-Knowledge', Presidential Address to the Pacific APA, *Proceedings and Addresses of the American Philosophical Association* 73:2 (November 1999), 25–46.

The account here also answers objections to my 'Individualism and Self-Knowledge', by Sven Bernecker, 'Externalism and the Attitudinal Component of Self-Knowledge', *Noûs* 30:2 (June 1996), 262–275. Bernecker noted that 'Individualism and Self-Knowledge' did not explain warrant in knowing the modes of our attitudes (belief rather than supposition, intention rather than hope). As I have emphasized elsewhere, 'Individualism and Self-Knowledge' was not intended as an epistemic account of Self-Knowledge, much less self-Knowledge.

apperceptive point of view that grounds applicability of norms of critical reason.[109]

VII

I now discuss a further aspect of betokening understanding. Critical reasoners must use understanding of lower-level reason-support relations to guide betokening understanding of instances of propositional attitudes. Betokening understanding must be guided not only by competence with lower-level preservation relations, but also by conceptualizations of reason-support relations among the target attitude and other lower-level attitudes. Such relations place the target attitude instance in a network of reason relations. (See note 108.) Any understanding of attitude instances that could not conceptualize reason-support relations as such could not ground the applicability of norms of morality and critical reason. In betokening an attitude instance, Self-Understanding essentially uses *de re* understanding of specific reason-support relations, practical or theoretical, among that attitude and others. Understanding an attitude's place in a network of reasons, for or against, bears on epistemic entitlement to judgments based on betokening understanding of that attitude. Since the identity of propositional attitudes constitutively depends partly on reason relations to other attitudes, betokening informed by understanding such connections is a better route to veridicality than any would-be betokening not so informed. So this identification through reason support in betokening understanding contributes to warrant for Self-Knowledge.

As with betokening understanding, lower-level competence understanding is used *in* the meta-representational understanding of reason-support relations as such (*R* is a reason for *A*). Understanding reason-support relations requires non-meta-representational competence to *use* one content as reason for another, and some disposition to be moved by, to *implement*, relations of threat or support. To understand a reason as such, for or against an attitude, one must have some lower-level disposition to shift or maintain attitudes in accord with the reason. Recognizing a reason as such essentially relies on the lower-level competence to use the reason to support the conclusion. This competence figures in the conceptual mode of presentation of the reason-support relation between a given betokened attitude and others.[110]

Take the concept, is a reason for believing that someone loved Haydn. Suppose it true of an instance of a belief that Mozart loved Haydn. The

[109] I presented this account in the oral form of these lectures in December 2007 at Columbia University. The general outlines of this account of one's warrant for Self-Knowledge are adumbrated independently by Frank Olav Barel in his Ph.D. dissertation 'Self-Knowledge and Anti-Individualism: A Defence of Neo-Rationalistic Compatibilism', (University of Oslo, 2010).

[110] Cf. 'Reason and the First Person'.

betokening understanding of that belief might be partly guided by an understanding of its supporting a belief that someone loved Haydn. Understanding the instance's supporting and being supported by other possible attitudes helps guide the betokening understanding. Concepts of reason support are meta-representational. In using them, however, a thinker must rely on a lower-level competence understanding of the *represented* propositional attitude types—belief that Mozart loved Haydn and belief that someone loved Haydn—including a lower-level competence to reason from one to the other. One relies on lower-level understanding of the represented contents as part of thinking about those contents, and about reason-support relations among them. The key point is that the guiding conceptualizations of reason-support relations employ lower-level competence understanding of the targeted contents, modes, and reason-support relations.

This point supplements, I think constitutively, the account already given of immunity to brute error yielded by the default warrant for betokening understanding and judgments based on it. Meta-representational understanding of reason-support relations must use lower-level competence understanding. The lower-level competence understanding is warranted through correctly representing the relations of reason support that set norms for epistemic warrant. The meta-representational understanding is warranted by relying on correct lower-level competence understanding of reason-support relations. So a warrant for meta-representational understanding of reason-support relations yields immunity to brute error.

To summarize the idea crudely, meta-representational understanding of reason-support relations depends for its warrant on sucking up warrant, present in the lower-level competence, for engaging in lower-level reasoning. To be warranted, the meta-representational understanding must involve correct understanding, and involve a disposition to be moved by, the lower-level reason-support relations. The meta-representational understanding of reason-support relations helps guide betokening understanding of attitude instances by locating them in a local network of practical and theoretical reasons. Meta-representational understanding of such relations improves one's route to veridical betokening understanding. Meta-representational understanding of reason-support relations contributes warrant to betokening understanding by helping it follow this route, only if it is correct understanding. It must be understanding of relations of genuine reason support. So both the warrant for meta-representational understanding of reason-support relations and the contribution of warrant for such understanding to the warrant for betokening understanding yield immunity to brute error. Clearly, this aspect of the warrant for betokening understanding is constitutive to critical reason—not merely a component of critical reasoning whose warrant comes from outside critical reason itself. (See note 108.)

VIII

I turn to betokening understanding in Self-Understanding and Self-Knowledge of two types of *non-propositional* psychological states and occurrences. One type comprises nonpropositional *representational* empirical states—such as perceptions and perceptually derived images. The other comprises conscious sensory states, such as certain aspects of pains, that are not representational in my sense. I will have to deal with these cases briefly.

The account of betokening understanding of perceptions, perceptual anticipations, perceptual memories, and fictional images broadly follows the account of betokening understanding of propositional states. Capacities to preserve mode and content of perceptions are constitutive to any empirical representational psychology. In a propositional-attitude psychology, such preservational connections obtain between perception and first-order perceptual belief. If a psychology has meta-representational capacities, some of them conceptualize lower-level contents *in* specifying those contents. Just as object-level perceptual concepts use perceptual contents in their modes of presenting environmental entities, so meta-concepts that represent perceptual contents—especially perceptual attributives—employ the object-level perceptual contents in *their* modes of representing those contents. Thinking about perceptions also employs lower-level mechanisms for preserving perceptual mode in conceiving of that mode, type and instance. Meta-representation that uses lower-level capacities is constitutive to the apperceptive core self.

An individual is default entitled to beliefs based on reliably veridical betokening understanding. Reliability is necessary for there to *be* the relevant point of view. Errors derive only from malfunction or misuse of the preservational operations—undermining entitlement—or from failure to base belief on those operations. A relevantly based judgment cannot be mistaken about a current perception, or immediately past perception, or a perceptual anticipation, while all is well with the cognitive apparatus. Brute error—warranted, representationally well-functioning error—is not possible relative to the relevant warrant.

Betokening understanding of *non-representational* conscious sensations, or aspects of them, differs importantly from the other types of betokening understanding. No lower-level capacities preserve representational content. But the basic shape of the account of the warrant for Self-Understanding of conscious sensations is otherwise similar.

Phenomenal consciousness of sensations is constitutive to being a self. Rational access to such consciousness is also constitutive.[111] Here betokening

[111] Higher animals and very young children that lack selves can surely form beliefs about sensations like pain. Such beliefs are as primitive as ordinary perceptual beliefs about the environment. Of course, such beings cannot *self*-attribute having those sensations; for they are not selves. But their beliefs make the sensations ego-relevant through whatever ego-concept they have that precedes a full-fledged self-concept. Traditional theology might allow that God is a self and lacks sensations. I confine my thesis to finite selves that interact with the physical world.

understanding forms concepts of sensations. The concepts constitutively use the sensations, or memories of their types, in representing them. So a lower-level element is again used in conceptual representation of it. A capacity to use such concepts *de re* to pick out instances is not constitutive to reasoning. But it is constitutive to being a self. A self must be able to think about and act on some conscious affect, such as its pains or pleasures. Such a capacity is basic to owning and controlling bedrock aspects of conscious life—the core of consciousness in a point of view. Being subject to moral norms also seems to hinge on this capacity.

A reliable natural capacity to form true beliefs in normal circumstances—those in which the beliefs' representational natures were formed—yields epistemic entitlement to those beliefs. Normally, not all is cognitively well with a person who mistakenly believes he or she is in sharp pain, or that a sharp pain is a tickle. Error can, of course, arise from anticipation, inference, or ideology. But any error that relies on the most direct conscious way of thinking about a sensation is a dislocation of self—a pathology that undermines warrant.[112] Beliefs about one's conscious sensations are immune to brute error, *relative to the entitlement*.[113]

This completes my sketch of applications of betokening understanding, and judgments based on it, to non-propositional cases.

IX

Predications in Self-Understanding are *self-attributions of* betokening understanding. Self-attribution is attribution of the psychological state to oneself via self-conception.

What is the warrant for the transition from warranted *de re* betokening understanding to self-attribution? Wherein is one entitled, in Self-Understanding, to attribute *to oneself*—referred to as oneself, with a self-concept—an instance of an understood state?

[112] There are borderline cases. Perhaps some pains are so close to borderlines with other types of sensation that, even in ordinary circumstances, one could make a brute error about whether one is having one. (Presumably, one could still get right that it is a sensation of a more generic type.) Perhaps one could even be *unreliable* in forming judgments about certain types of sensations that are easily mis-categorized, at least given only short periods of time. Such cases do not, I think, count against the view that many judgments about conscious sensations are immune to brute error relative to certain warrants associated with relevant judgment-forming routes.

Again, I think that whether a warrant yields immunity to brute error is not apriori determinable from the content of the belief. *A fortiori*, I think that immunity to all error is not intrinsic to self-attributions with sensation concepts. I think that because of cognitive malfunction or stoic ideology, an individual can judge even a very intense pain not to be a pain.

[113] I think that *some* judgments about *past* sensations, and perceptions, are immune to brute error relative to certain warrants. The relevant warrants attach to judgments that use purely preservative memory of singular applications of concepts of those sensations or perceptions. Here the account parallels that of betokening understanding of past propositional thought events.

The transition is idealized. There is no step in time. Betokening understanding and self-attribution are aspects of the predication. The relation between the aspects is, however, usefully considered as a transition step, although there is no literal inference. It is *not* an inference from That is a belief that *p* to I believe that *p* (or that belief is mine).

The epistemic norm for the transition step, governing warrant for basic apperception is as follows:

> (WBA) If one is entitled in the relevant way to betokening understanding of an instance of a psychological element, and if the self-attribution relies only on the betokening understanding and on one's competence with the self-concept, one is entitled to attribute that instance to oneself as oneself.

This norm is grounded in the fact that what one is entitled to believe about one's self rests fundamentally on what one is entitled to believe about one's psychological states. Selves are not systems of psychological states. They are the subjects/agents of such systems. But the natures of selves consist in their psychological competencies. As noted in the first lecture, the capacity to form warranted beliefs about one's self through beliefs about one's psychological states has its phylogenetic prehistory in couplings between occurrences of *de se*, or ego-centric, indexes applied to an individual as origin of a representational framework and as subject matter of perception. Such couplings center on perceivable bodily attributes. Selves are not perceivable. But we can know our psychological states *de re*. Such knowledge is a basic route to knowing our selves. Betokening understanding of instances of psychological elements grounds the *Self*-Knowledge that I have been discussing.

A central instance of self-cognition through cognition of one's psychological states is self-tracking in memory. Again, the relevant memory is grounded in more primitive couplings. *De se*, ego-centric indexes, the primitive ancestors of self-concepts, are embedded in all experiential memory.[114] When occurrences of such indexes *that mark an individual* are coupled with occurrences applied to a subject matter, they allow diachronic ego-tracking. The initial tracking is perceptual and body centered. Memories associated with couplings in passing the mirror test are likely examples. When memories become meta-psychological, indexes in them can be used in couplings that track past psychological instances, for example, transitions in reasoning, in meta-psychological autobiographical episodic memory. Such trackings are among the diachronic unities that make selves possible.

These unities are grounded in present-tense couplings—those that attribute current psychological instances, identified through betokening understanding, to a subject/agent that is a central *locus* for them. Descartes' *cogito* focused on performative present-tense couplings. But there are equally important couplings

[114] For more on primitive ancestors of the self-concept, see my 'Memory and Persons'; and '*De Se* Preservation and Personal Identity: Reply to Shoemaker', this volume.

with present standing states. When couplings of ego-centric indexes are incorporated into betokening understanding of psychological states and used in critical reasoning, such indexes support self-concepts. Having a self-concept, indeed being a self, is grounded in reflexive psychological competencies that meet epistemic norms for Self-Understanding.

Fulfilling these epistemic norms is making connections constitutive to selves. Making the connections—engaging in the diachronic and multi-tiered preservations—just is getting things right about one's psychology and one's self. Thus fulfilling the norm set out in (WBA) yields immunity to brute error for self-attributions. The step's starting point—warranted betokening understanding—cannot introduce brute error. If self-attribution relies on that understanding, it veridically self-attributes the Self-Understood state. Self-attributions based in betokening understanding are paradigmatically warranted uses of the self-concept. Errors in self-attribution derive from warrant-undermining pathology or from not relying on the relevant preservational mechanisms. So self-attributions are immune to brute error, relative to the relevant warrants.

Conceiving the psychological elements as one's own via the self-concept is acknowledging ownership of commitments in the psychological elements. Capacity for such acknowledgment helps ground applicability of norms of critical reason and morality.

Many issues arise about this account of warrant for self-attribution. I mention two.

Shoemaker's discussion of quasi-memory may seem to threaten immunity to brute error in Self-Understanding. Through quasi-memory, an individual could identify an attitude instance that is *not* the individual's own in a *de re* way. The individual could make a brute error in self-attributing the quasi-remembered psychological state. Remembering that warrants attach to psychological abilities helps disarm such threats. An ability to preserve one's own attitudes differs, psychologically and epistemically, from any ability to tap into other minds. Brute errors of self-attribution relative to warrants associated with quasi-memory are irrelevant to self-attributions warranted by memory.[115]

Some schizophrenics think that their thoughts are inserted into their minds by others. They deny that their thoughts, which they cite correctly, are their own. Such cases do not threaten the account because they involve pathology in the preservative capacities that ground entitlements to self-attribution. The schizophrenics do not rely purely on betokening understanding to make a self-attribution, in the way specified by the norm. There is some malfunction in connecting self-ascription with betokening understanding. Such malfunction undermines warrant for self-ascription in judgments about state instances. The

[115] Shoemaker, 'Persons and Their Pasts', *American Philosophical Quarterly* 7:4 (October 1970), 269–285, reprinted in his *Identity, Cause, and Mind* (New York: Oxford University Press, 2004). See also my 'Memory and Persons'.

errors are not errors in exercising well-functioning capacities that constitute apperceptive points of view.[116]

My account of warrant for Self-Understanding and Self-Knowledge refines Kant's notion of unity of apperception. Unity lies in an understanding that utilizes transtemporal and inter-level constitutive connections in the structure of selves.

The account highlights a *singular intellectual* power, *de re* understanding of particulars. Descartes emphasized *de re* understanding in contrast to Aquinas's emphasis on reason as a competence with generality. Reason and *de re* intellectual understanding are deeply interrelated. But they are distinct. Even now, the latter is neglected in philosophy.

Kant took up Descartes' notion in his own conception of understanding of particulars. Kant insisted that such understanding must be warranted through sense capacities. I think the insistence mistaken. I think that Descartes was right to hold that we have an understanding of particulars that has a purely intellectual warrant. Descartes over-rated the capacity's power for metaphysical insight. But he was right to highlight the capacity. Self-Understanding provides a central example. Other examples occur in mathematics, certain types of appreciation of intellectual beauty, and perhaps moral thinking.

I hope to have delineated the centrality of Self-Understanding in being a self—and in having the cognitive and valuational capacities, and being subject to the norms, that make us special. Using our intellects to understand the particular, particularly ourselves, is a way in which we are valuable, minded islands in a largely mindless, value-neutral universe.

[116] See Christopher D. Frith, *The Cognitive Neuropsychology of Schizophrenia* (Hove, UK: Lawrence Erlbaum, 1992); Simon Mullins and Sean A. Spence, 'Re-Examining Thought Insertion: Semi-Structured Literature Review and Conceptual Analysis', *The British Journal of Psychiatry* 182 (2003), 293–298.