

Wiser, M., and S. Carey. (1983). When heat and temperature were one. In D. Gentner and A. Stevens, Eds., *Mental Models*. Hillsdale, NJ: Erlbaum, pp. 267–298.

Further Readings

- Giere, R. N. (1992). Cognitive models of science. *Minnesota Studies in the Philosophy of Science* 15: 3–44.
- Klahr, D. (1994). Searching for cognition in cognitive models of science. *Psychology* 5(69). 94.5.69.
- Shrager, J., and P. Langley. (1990). *Computational Models of Scientific Discovery and Theory Formation*. San Mateo, CA: Kaufmann.
- Tweney, R. D., and S. Chitwood. (1995). Scientific reasoning. In S. E. Newstead and J. St. B. T. Evans, Eds., *Perspectives on Thinking and Reasoning: Essays in Honour of Peter Wason*. Hove, England: Erlbaum, pp. 241–260.
- Valdes-Perez, R., and D. H. Sleeman, Eds. (1997). *Artificial Intelligence* 91(2), special issue on Scientific Discovery.

Scripts

See SCHEMATA

Search

See GREEDY LOCAL SEARCH; HEURISTIC SEARCH

Self

Questions about the self are typically posed as questions about persons or minds, about such self-reflexive capacities as SELF-KNOWLEDGE and self-reference, or about the semantics and pragmatics of “I.” For example, we think of the self or person as something that endures through changes in its mental states; but what is it that makes us the same person now who we were ten years ago? Among those who reject the idea of a nonphysical substance or soul, the debate has focused on the relative importance of bodily continuity (especially continuity of the brain) and psychological continuity (Williams 1973; Parfit 1984). Because the focus on psychological continuity entails that there could in principle be more than one person in a single human body, the debate has clear implications for controversies in clinical psychology such as that surrounding multiple personality subjects (Hacking 1995). More recently the debate has expanded to include such normative issues as the nature of the justification of the sacrifices that we ordinarily make for our future selves (White 1991; Rovane 1997), raising the question whether personhood is a metaphysical or a normative concept.

Besides thinking of ourselves as enduring, we have the idea of ourselves as agents—as subjects of actions and not merely objects to which things happen. This raises the issue of whether we could make sense of agency, freedom, and responsibility if all our actions were causally determined or, indeed, whether we could do so even if they were uncaused and random. Compatibilists hold that free will is not a matter of our actions being uncaused but of their being caused

in the right way—for example, by a process of deliberation that is uncoerced, uncompelled, and so forth (Ayer 1954). But this leaves open the question how we could be justified in allocating and accepting responsibility for actions that were determined to happen long before we were born (Strawson 1962; White 1991). Another issue that compatibilism leaves unresolved is over the nature of the experience that grounds our concepts of freedom and agency. This topic has been addressed extensively in the existentialist and phenomenological traditions and is currently under investigation in psychology (Heidegger 1962; Sartre 1956; Merleau-Ponty 1962; Neisser 1993; Introduction).

Our concept of the self, however, is not simply that of an enduring entity to which certain mental events and actions are ascribed. We normally assume that our knowledge of ourselves as subjects is nonobservational and noninferential and thus unlike our knowledge of ordinary physical objects. One reason is that our thoughts about ourselves as subjects (thoughts expressed in terms of “I”) seem to enjoy an immunity to error that those regarding external objects lack. Although one might be mistaken about whether a body that one could observe was one’s own, there is no possibility of losing track of oneself as a subject or of mistaking another subject for oneself or another mind for one’s own (Shoemaker 1968). But though our knowledge of ourselves as subjects is apparently not a matter of external observation, it seems that anything to which we could have introspective access would be a mental state and not the enduring subject that has that state. David HUME (1888) concluded on this basis that we have no access to an enduring self and indeed that none exists, a position subsequently taken up by Ernst Mach (1939) and Moritz Schlick (1949).

Self-reference raises some of the same problems as self-knowledge and provides similar reasons for skepticism about the self. How does the term “I” refer? Evidently not through any associated linguistic descriptions, in that we can imagine experiencing amnesia while anesthetized in a sensory deprivation chamber and thinking “I won’t let this happen again!” (Anscombe 1975). In such a situation we could not frame a description that would pick us out uniquely. Nor is it adequate to say simply that “I” refers demonstratively, because we normally perceive the object that we demonstrate. Anscombe (1975) has argued on the basis of considerations of this kind that “I” does not refer.

But neither the claim that “I” does not refer nor its denial answers the most basic question underlying the issues of self-reference and self-knowledge and the most basic question associated with the self—How *are* we given to ourselves when our access to ourselves is most immediate? One approach to this question is suggested by work in phenomenology and by contemporary psychologists influenced by James Jerome GIBSON (Heidegger 1962; Sartre 1956; Gibson 1986; Neisser 1988). These theorists hold that what we are given is a pragmatically structured world of opportunities and liabilities. These are perceived directly and immediately; they are not interpretations imposed on a neutral sensory field the perception of which is more direct or immediate. Gibson’s term for what is given in such experiences is AFFORDANCES.

As both Sartre and Gibson make clear, in being given a world of human possibilities and things to be done—of doorways that we can walk through and streetcars we can catch if we hurry—we are given ourselves *implicitly*. A world of affordances is one in which we are necessarily implicated. To see the chair as something to sit on is to be given to ourselves (implicitly) as having a certain size, weight, and shape and a certain capacity for movement and action. And our perceiving the speeding car as a threat says as much about our vulnerabilities and liabilities to destruction as it does about automobiles (Warren and Whang 1987; Mark 1987).

The notion of a self implicit in our perceptual experience of the external world raises the question how such a self is related to our explicit conception of ourselves as objective entities. Strawson (1959) poses a similar question by asking how we can have a conception of an entity to which we ascribe both mental and physical properties. And the question how we can ascribe both mental and physical properties to ourselves raises many of the same issues as the question how we can ascribe the same mental properties to ourselves and to others. The problem is that our basis for the ascription of mental properties to others (observation of behavior) is so radically different from our basis for self-ascription that a commitment to the idea of meaning as use suggests that mental predicates must change their meanings from first-person to third-person contexts. This, in other words, is one version of the problem of other minds.

If views like those of Gibson and the phenomenologists that take the agential perspective as basic are correct, however, this way of posing the problem may exaggerate the asymmetries between ourselves and others. Recent research suggests that our capacity to engage in joint or complementary actions with others is in place at birth in the form of a capacity to engage in and recognize mimicry and slightly later in the ability to share affect in expressive exchanges (Meltzoff and Moore 1995; Stern 1985). This points to the possibility that our access to the other subjects' agential characteristics may be at least as direct as our access to their objective makeup. For example, just as we are given ourselves implicitly in the possibilities we see for individual action, the possibilities we see for acting jointly may give us an implicit other—a notion similar to Sartre's (1956: part III, chap. 1) understanding of Heidegger's "being-with." Alternatively, what we perceive most immediately may be a relation of INTERSUBJECTIVITY—"an appropriate match between the nature/direction/timing/intensities of two people's activities" (Neisser 1988: 41). If this general approach can be sustained, then the question how we can ascribe mental properties to an objectively characterized other is misleading. From the agential perspective the problem is rather one of acquiring a more objective conception both of one's partner and of oneself. And, as a large body of contemporary work in psychology suggests (Neisser 1993; Cicchetti and Beeghly 1990; Butterworth 1982), this reformulation may prove more tractable than the original problem.

See also ECOLOGICAL PSYCHOLOGY; IMITATION; SIMULATION VS. THEORY-THEORY; SITUATEDNESS/EMBEDDEDNESS

—Stephen L. White

References

- Anscombe, G. E. M. (1975). The first person. In S. Guttenplan, Ed., *Mind and Language: Wolfson College Lectures*. New York: Oxford University Press.
- Ayer, A. J. (1954). Freedom and necessity. In A. J. Ayer, Ed., *Philosophical Essays*. London: Macmillan.
- Butterworth, G. E., Ed. (1982). *Infancy and Epistemology: An Evaluation of Piaget's Theory*. New York: St. Martin's Press.
- Cicchetti, D., and M. Beeghly, Eds. (1990). *The Self in Transition: Infancy to Childhood*. Chicago: University of Chicago Press.
- Gibson, J. J. (1986). *The Ecological Approach to Visual Perception*. Hillsdale, NJ: Erlbaum.
- Hacking, I. (1995). *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton: Princeton University Press.
- Heidegger, M. (1962). *Being and Time*. New York: Harper and Row.
- Hume, D. (1888). *A Treatise of Human Nature*. L. A. Selby-Bigge, Ed. Oxford: Clarendon Press.
- Mach, E. (1939). *The Analysis of Sensations*. New York: Dover.
- Mark, L. S. (1987). Eyeheight-scaled information about affordances: A study of sitting and stair climbing. *Journal of Experimental Psychology: Human Perception and Performance* 13: 361–370.
- Meltzoff, A. N., and M. K. Moore. (1995). Infants' understanding of people and things: From body imitation to folk psychology. In J. L. Bermúdez, A. Marcel, and N. Eilan, Eds., *The Body and the Self*. Cambridge, MA: Bradford/MIT Press.
- Merleau-Ponty, M. (1962). *Phenomenology of Perception*. London: Routledge and Kegan Paul.
- Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology* 1: 35–59.
- Neisser, U., Ed. (1993). *The Perceived Self: Ecological and Interpersonal Sources of Self-Knowledge*. New York: Cambridge University Press.
- Parfit, D. (1984). *Reasons and Persons*. New York: Oxford University Press.
- Rovane, C. (1997). *The Bounds of Agency: An Essay in Revisionary Metaphysics*. Princeton: Princeton University Press.
- Sartre, J.-P. (1956). *Being and Nothingness*. New York: Simon and Schuster.
- Schlick, M. (1949). Meaning and verification. In H. Feigl and W. Sellars, Eds., *Readings in Philosophical Analysis*. New York: Appleton-Century-Crofts.
- Shoemaker, S. (1968). Self-reference and self-awareness. *Journal of Philosophy* 65: 555–567.
- Stern, D. N. (1985). *The Interpersonal World of the Infant*. New York: Basic Books.
- Strawson, P. F. (1959). *Individuals*. Garden City, NY: Doubleday.
- Strawson, P. F. (1962). Freedom and resentment. *Proceedings of the British Academy* 48: 1–25.
- Warren, W. H., and S. Whang. (1987). Visual guidance of walking through apertures: Body-scaled information for affordances. *Journal of Experimental Psychology: Human Perception and Performance* 13: 371–383.
- White, S. L. (1991). *The Unity of the Self*. Cambridge, MA: Bradford/MIT Press.
- Williams, B. (1973). *Problems of the Self*. New York: Cambridge University Press.

Further Readings

- Bermúdez, J. L., A. Marcel, and N. Eilan, Eds. (1995). *The Body and the Self*. Cambridge, MA: Bradford/MIT Press.
- Campbell, J. (1994). *Past, Space, and Self*. Cambridge, MA: Bradford/MIT Press.

- Cassam, Q., Ed. (1994). *Self-Knowledge*. New York: Oxford University Press.
- Cassam, Q. (1997). *Self and World*. Oxford: Clarendon Press.
- Dennett, D. C. (1992). The self as a center of narrative gravity. In F. Kessel, P. Cole, and D. Johnson, Eds., *Self and Consciousness: Multiple Perspectives*. Hillsdale, NJ: Erlbaum.
- Eilan, N., R. McCarthy, and B. Brewer, Eds. (1993). *Spatial Representation*. Oxford: Blackwell.
- Elster, J. (1986). *The Multiple Self*. New York: Cambridge University Press.
- Evans, G. (1982). *The Varieties of Reference*. New York: Oxford University Press.
- Gallagher, S., and J. Shear, Eds. (1997). Special issue: Models of the self. *Journal of Consciousness Studies* 4: 385–540.
- Husserl, E. (1950). *Cartesian Meditations*. Dordrecht: Kluwer.
- James, W. (1950). The consciousness of self. In *The Principles of Psychology*. New York: Dover.
- Lewis, M., and J. Brooks-Gunn. (1979). *Social Cognition and the Acquisition of Self*. New York: Plenum.
- Mead, G. H. (1934). *Mind, Self, and Society*. Chicago: University of Chicago Press.
- Peacocke, C., Ed. (1994). *Objectivity, Simulation, and the Unity of Consciousness*. New York: Oxford University Press.
- Radden, J. (1996). *Divided Minds and Successive Selves: Ethical Issues in Disorders of Identity and Personality*. Cambridge, MA: Bradford/MIT Press.
- Sartre, J.-P. (1957). *The Transcendence of the Ego: An Existentialist Theory of Consciousness*. New York: Farrar, Straus, and Giroux.
- Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*. Ithaca: Cornell University Press.
- Taylor, C. (1989). *Sources of the Self: The Making of the Modern Identity*. Cambridge, MA: Harvard University Press.
- Tugendhat, E. (1986). *Self-Consciousness and Self-Determination*. Cambridge, MA: MIT Press.
- Vygotsky, L. (1986). *Thought and Language*. Cambridge, MA: MIT Press.
- Wilkes, K. (1988). *Real People: Personal Identity without Thought Experiments*. New York: Oxford University Press.
- Wittgenstein, L. (1958). *The Blue and Brown Books*. Oxford: Blackwell.

Self-Knowledge

The beginning Socratic injunction “Know thyself” dates to the very beginning of the Western intellectual tradition, and self-knowledge in its many forms has been a central concern ever since. A wide variety of cognitive states fall under its conceptual umbrella. The sort of self-knowledge acquired through cognitive psychotherapy, for example, might include explicit beliefs about the motives of one’s behavior couched in the concepts of folk psychological intentional discourse and available for application in the control or modification of future personal action. Alternatively, the Chomskyan self-knowledge that humans are alleged to have of the structure of their language processing systems is largely implicit, subpersonal, innate, and limited in application to the specific process of language-learning.

The types of self-knowledge vary along four main parameters:

- Content
- Manner or mode of representation

- Domain of application
- Means of acquisition

Although the four are importantly interdependent, each presents its own set of issues.

The content of self-knowledge essentially concerns what a cognitive agent knows about its own nature or organization. Though paradigm cases involve knowledge about one’s psychological nature, any property or feature of the agent can be its object. One can have self-knowledge of the size, state, and orientation of one’s body as well as of one’s beliefs and emotions. Despite this breadth, it is the cases of metapsychological cognition that are of greatest interest; for it is they that provide the opportunity for radical increases in mental sophistication. It is virtually impossible to create a system with highly sophisticated mental abilities without building in a significant degree of metapsychological understanding. This point can be overlooked if one focuses exclusively on cases of explicit human self-knowledge of the sort that could be verbally reported. However, if one recognizes the wealth of implicit self-knowledge that must be implicitly embodied in the systems of self-monitoring and self-regulation required by any sophisticated cognitive agent, it becomes readily apparent that self-knowledge of one sort or another will be the pervasive and central feature of all but the most minimally minded systems. Even simple organisms and robots require some measure of self-knowledge to function as cognitive actors. Without knowledge of one’s goals and abilities, it would be impossible to carry out multiple stage actions. And no learning would be possible in a system that totally lacked understanding of the function and organization of the processes to be modified. The requisite self-knowledge might all be implicitly embedded in the structure of the learning mechanisms, but that organization in itself counts as a form of self-knowledge insofar as it adaptively reflects the nature of the processes that are changed through learning.

Explicit human self-knowledge nonetheless remains of great interest to philosophers and psychologists. Adult humans have unique abilities to reason about their mental lives and generally reliable insight into their inferential processes, motives, and preferences. This enhances their ability to regulate their mental lives and to interact with others in social contexts. Empirical studies have nonetheless shown glaring gaps in human metacognitive powers; in specific test situations adult subjects show ignorance of the factors governing their choices, their sources of information, and the rules underlying their reasoning.

The degree to which explicit self-knowledge may be present in children or in nonhuman primates is controversial. Children under the age of three fail to distinguish between their own beliefs about a situation and those that an agent with different access would have; they seem unable to conceptualize their own view of reality as such, that is, as one among many possible views. Tests, especially those involving deceptions, on chimpanzees seem to show some grasp of explicit mental concepts, but only of the most limited sort. Other studies, such as those using mirror recognition tasks, give evidence that chimps and orangs have some type of self concept, but again of only a limited sort. Further research is needed to resolve these issues. In a surprising

twist, the psychologist Alan Leslie has proposed that an inability to engage in METACOGNITION is the primary deficit in AUTISM and the source of most of the disabling symptoms associated with it.

Cases of self-knowledge vary not only in their objects but also in how they conceptualize or categorize their objects. Explicit human self-knowledge of the aims of one's behavior is likely to be conceptualized in the folk psychological notions of belief, desire, and intention, but such concepts are not likely to figure in the self-knowledge one's retrieval system has of the structure and organization of one's memory nor in a cat's knowledge of its current needs and goals. Indeed, there may be no public-language words or concepts that adequately capture the content of such implicit knowledge. But insofar as it partitions its psychological objects in similarity classes and generates aptly matched responses, even such procedural self-knowledge must involve categorization or conceptualization.

Self-knowledge can be explicitly represented and stored as is likely the case with the propositional knowledge that humans have of their intentions and beliefs. The psychologist Philip Johnson-Laird has argued that the planning and control of intentional action requires a self model that explicitly represents one's goals, abilities, options, and current state. But significant self-knowledge can also be implicitly represented in the structure of a network or in the organization of a metacontrol system as is probably true of that embodied in many learning processes. The mechanism by which rats learn to avoid foods that have been followed by bouts of nausea several hours after feeding provides an apt example. One need not suppose that the rat has any explicit awareness of the processes that regulate its feeding behavior or taste sensations, yet its nervous system clearly carries information about those factors as shown by its ability to alter those mechanisms in just the way needed to produce the desired behavioral change.

Procedurally embodied self-knowledge will often be more limited in its scope of application; the metapsychological understanding carried in a learning process may have no impact outside the context of the specific modifications it is designed to produce. The rat can not reflect on the organization of its feeding mechanisms nor make open-ended use of that information, as one might if one had explicit propositional knowledge of them. The difference, however, is one of degree. Procedural self-knowledge can have a relatively broad range of application, and even explicit beliefs about one's mental state are limited in their impact to some degree by the larger context within which they occur.

Self-knowledge can arise in many ways. Traditional Cartesian mentalism treated the mind as fully transparent and open in all its significant properties to a faculty of conscious INTROSPECTION or reflection, which was conceived of by later empiricists as a form of inner perception. Though introspection is now regarded as fallible, incomplete, and theory-laden, it nonetheless remains a major source of self-knowledge. Though internal monitoring, like external monitoring through the senses, is subject to error, it still provides a regular ongoing supply of information about the current state and operation of at least some aspects of one's mind.

The development of self-knowledge depends on both cultural and biological sources. Some theorists argue that the child's mastery of mental concepts involves building a behaviorally based theory of mind; others see it more as a matter of projection from first person-based concepts. In either case, both innate information and culturally based learning will be involved. Children must acquire from their social context many of the concepts needed to categorize their mental states and processes, but folk psychology is likely also to embody an innate scheme of mental categories. Moreover, the neural mechanisms that underlie introspection probably depend on innate implicit self-knowledge in the same way that our perceptual processes depend on such knowledge about the environment.

Though self-knowledge has many obvious benefits, it may not always be adaptive. Ignorance of one's limitations may enhance one's ability to mobilize oneself to action, and a lack of self-knowledge may signal a happy freedom from narcissistic self-absorption. This creates a new puzzle about self-deception. The old puzzle was to explain how deceiving oneself was even possible, given the apparently self-defeating identity of deceiver and deceived. If self-deception is often adaptive, a new more normative problem arises. Must we abandon our intuition that self-deception in itself is a bad thing? Must we qualify the Socratic "Know thyself" with the proviso "but only when it's useful"?

See also EPISTEMOLOGY AND COGNITION; INNATENESS OF LANGUAGE; MENTAL MODELS; PRIMATE COGNITION; SELF

—Robert van Gulick

References

- Beckoff, M., and D. Jameison. (1996). *Readings in Animal Cognition*. Cambridge, MA: MIT Press.
- Bennett, J. (1988). Thoughtful brutes. *Proceedings and Addresses of the American Philosophical Association* 62: 196–211.
- Chomsky, N. (1980). *Reflections on Language*. New York: Pantheon.
- Garcia, J. A., and R. A. Koelling. (1967). A comparison of aversions caused by x-rays, toxins and drugs in rats. *Radiation Research Supplement* 7: 439–450.
- Goldman, A. (1993). The psychology of folk psychology. *The Behavioral and Brain Sciences* 16: 15–28.
- Gopnik, A. (1993). How we know our minds: The illusion of first-person knowledge of intentional states. *Behavioral and Brain Sciences* 16: 15–28.
- Johnson, M., S. Hashtroudi, and D. Lindsay. (1993). Source monitoring. *Psychological Bulletin* 114: 3–28.
- Johnson-Laird, P. (1983). *Mental Models*. Cambridge, MA: Harvard University Press.
- Lyons, W. (1986). *The Disappearance of Introspection*. Cambridge, MA: MIT Press.
- McLaughlin, B., and A. Rorty. (1988). *Perspectives on Self-Deception*. Berkeley: University of California Press.
- Nisbet, R., and T. Wilson. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84: 231–259.
- Rorty, A. (1975). Adaptivity and self-knowledge. *Inquiry* 18: 1–22.
- Wellman, H. (1990). *The Child's Theory of Mind*. Cambridge, MA: MIT Press.