

## Models of the self

SHAUN GALLAGHER & JONATHAN SHEAR  
(Eds)

Thorverton: Imprint Academic, 1999

ISBN 0907845401 (hardcover)

ISBN 0907845096 (paperback)

*Models of the Self*, collects 27 essays from four volumes of the *Journal of Consciousness Studies*. The explicit aim of the collection is to "explore various dimensions of the self by drawing on a diverse set of disciplines and approaches" (p. xi). Included are the perspectives of analytic philosophy, developmental and evolutionary biology, transcendental meditation, neurology, neuroscience, phenomenology, developmental and cognitive psychology, psychiatry and robotics. The collection is uneven in quality and, taken as a whole, amounts to what Galen Strawson calls a "festival of misunderstanding" (p. 483). Yet sometimes disorienting and confusing festivities are the most fun (even if they are unlikely to leave you feeling clear headed), and perhaps the study of the self is too young to abandon such raucous ways.

The chaperones to this affair (a 10-page introduction by Gallagher & Shear and two articles by Strawson) do little to maintain order. The introduction (Gallagher & Shear) consists of a two-page romp through the history of philosophy from Aristotle to Hume (a curious terminus) and a series of paragraph-length chapter summaries. As such, it does very little to identify or diffuse misunderstandings, to highlight overlap or to stimulate further communication between these different disciplinary perspectives. This task is left to two bookend pieces by Strawson: "The self" (the first essay of the collection) and "The self and the SESMET" (a "response" article that is the last). This editorial choice gives the misleading impression that the collection is organized around Strawson. In fact, his provocative opener receives far more mention than focused treatment in the other essays. Consequently, Strawson's response essay addresses only those that seriously engage him, addresses them only to the extent that they engage him, and so does little to keep things from getting out of control.

The guests are loosely clumped into six groups: (i) philosophical controversies, (ii) cognitive and neuroscientific models, (iii) developmental and phenomenological constraints, (iv) pathologies of the self, (v) meditation-based ap-

proaches, and (vi) further methodological questions. Not all of these subsections are internally coherent (for example, (iii) is a composite of two unrelated issues and (vi) is a catch-all), and the distinctions between them are at times unnatural (there is considerable overlap between (i) and (vi) and between (ii) and (iv)). Few will read this book from cover to cover. I, for one, could muster only a passing interest in (v) but found (ii) and (iv) to be generally quite informative. Still, anyone interested in some question about the self will find something worth reading. Since this is not a party of the usual suspects, a few choice introductions will help to break the ice.

E. Olson argues that a volume like *Models* is likely to end in confusion. The reason: there is no single nontrivial account of the "problem of the self." There is no agreed upon definition of the self, no characteristic features that all notions of the self share, and no paradigm cases of selves to refer to in anchoring such a discussion. Different discussants take the self to be a subject of conscious experience, a person, a private inner being, what we value most, a mechanism of the unity of consciousness, a psychological attribute, or a bundle of sensations. (To this list we might add: an agent, a living organism, and a thread of diachronic identity.) Olson suggests that discussions carried out as if they were about the self are best recast as about other things, such as personal identity, the semantics of first-person pronouns, the unity of consciousness and knowledge of our own mental states. Although I suspect that Olson's pessimism glosses over significant interdependencies among these topics—interdependencies that likely explain the desire to speak of the self as their common subject—Olson nonetheless points the way to a more productive discussion. Those engaged in studies of the self should clearly identify the objects of and the problems addressed by their investigation, and Olson's taxonomy (like Neisser's (1994) five kinds of self-knowledge) is a helpful step towards productive communication.

V.S. Ramachandran and W. Hirstein (RH) are concerned with qualia (the qualitative aspect of conscious experience). They connect this concern to the self on the grounds that "the problem of qualia and the problem of the self are really just two sides of the same coin" (p. 84). This claim reveals an implicit conception of the self as the subject of conscious experience. Part I of the essay is advanced to "demolish" the philosophical work on qualia by Jackson, Kripke, Nagel and Searle (in

three pages, no less). Those who have thought seriously about the diverse positions defended by Jackson and the others will puzzle (but not for long) at RH's suggestion that all such philosophical walnuts might be cracked by a thought experiment involving building a neuronal bridge between the brains of neuroscientists and creatures with exotic qualia (specifically, human-like Amazonian electric fish). Philosophy is not the strength of this article. Part II, however, contains the unique blend of intriguing science and intellectual bravado that could allow Ramachandran to rival Oliver Sacks in popular appeal. RH introduce what they call "three laws of qualia" and defend those by reference to various visual illusions (e.g. the Müller-Lyer illusion) and clinical cases (e.g. Bonner's syndrome and phantom limbs). The three laws describe three functional characteristics that all qualia share: (i) they are *irrevocable* (i.e. unlike beliefs, qualia cannot be easily shed once we experience them), (ii) they give rise to *flexible outputs*, and (iii) they are *represented in working memory*, and so are available for making choices. What has all of this got to do with the self? RH argue that the self is an "illusion" sustained by the activities of a brain system that combines sensory qualia with affective and motivational states to "drive" planning and motor output. Their target here (the illusion) appears to be the same as Dennett's in his *Consciousness explained* (1991): a Cartesian theater or "inner person" who sees all of our experiences. It is not clear that any philosopher or scientist seriously endorses such a view anymore; but RH provide with this system an interesting explanation of why such a thought might have to be exorcised through careful scientific and philosophical work.

Unlike RH, Strawson believes that the self is a real, physical thing. He dubs his positive account of the self, "the pearl view." According to this view, our lives are composed of many selves, some lasting for only three seconds through "hiatus-free" periods of consciousness. These "selves" exist "one at a time and one after another, like pearls on a string" (p. 20), but without the string. Strawson arrives at this counterintuitive view through a "phenomenological investigation" of the sense of mental self (SMS). He argues from the phenomenal aspects of self-consciousness to the characteristics that the self must have to sustain the phenomenology. By "phenomenology" Strawson means something like "reasoned introspection," to the consternation of more traditional

"phenomenologists" such as M. Sheets-Johnstone (pp. 232–237) and of D. Zahzavi and J. Parnas (pp. 253–270). Strawson begins with his "local phenomenological question," which concerns the nature of the human sense of self. He claims, without argument, that the following SMS is culturally universal: the self is experienced as (1) a thing, (2) mental, (3) synchronically single, (4) diachronically single, (5) ontically distinct from other things, (6) a subject of experience, (7) an agent, and (8) a thing with a personality. He then turns to the "general phenomenological question," which asks after a bare minimum of phenomenological features for an SMS to count as a genuine SMS. Strawson concludes that (4), (7) and (8) are unnecessary for a minimal SMS or for selves. The road to Strawson's pearl view is bumpy and fun; it is dotted with interesting thought experiments and it is ambitious in its attempt to rethink the target for investigations of the self. Nonetheless, I suspect that many readers, like myself, will balk at both Strawson's phenomenology and his intuitions and so will be inclined to agree with Eric Olson: "Anyone who reaches the conclusion that none of us exist, or at least none for longer than a few seconds, must have gone wrong somewhere, just as certainly as someone who concludes that motion is impossible" (p. 58).

K. Wilkes criticizes Strawson for at once reifying the notion of the self and stripping it of the qualities that could make it explanatorily useful. First, Wilkes argues that Strawson is guilty of reifying "airy nothings" and that the purposes played by the concept of self are better played, as suggested by P.F. Strawson, by the concept of a "person" (a living human animal with diverse capacities). Second, Wilkes provides a useful reminder that the notion of a self is bound up with far more than can be found in moments of introspective self-consciousness. Wilkes argues that many human emotions (such as anticipation, contrition, dread, gratitude, hate, love, remorse and respect) and moral abilities (such as accepting praise and blame, calculating consequences, developing a character, overcoming defeat, and planning future actions) presuppose that people (selves) have duration. Wilkes argues further that there is considerably more to the self (or a person) than is available in our brief, hiatus-free moments of consciousness. Implicit beliefs, subconscious and subliminal determinants of actions, and unconscious desires are important aspects of our selves, but they do not so much as come into view

on Strawson's account. Brief and fleeting selves are hardly worth having. They cannot fulfill the explanatory work that the notion of a self might be thought to perform, and so there is no thing corresponding to Strawson's pearls.

Strawson looks within and stumbles on a self. J.L. Bermudez looks within and finds nothing but mental and physical states. In a short and difficult piece, Bermudez aims to clear space for a form of Humean reductionism about the self. Bermudez attempts to meet the challenge to reductionism (attributed to Campbell, 1994) to explain the legitimacy of first-person inferences. A first-person inference is an inference "from any two premises, both stated using the first-person" (e.g. I am F) to a conclusion stating "the identity of the thing referred to in the two premises—namely, himself" (p. 434). An antireductionist can ground such inferences in the sameness of the person (or self) referred to in the two premises. Bermudez eschews this approach, and attempts instead to ground the inferences in the fact that the premises refer to the same bundles of suitably connected mental and physical states (henceforth, bundles). The challenge for Bermudez, as it was for Hume, is to find a satisfactory account of the unity of such bundles. Bermudez responds to this challenge with a dilemma for the person-centered approach and a positive account. The dilemma (see p. 436) is as follows: either persons are something over and above bundles, in which case there are no clear criteria for individuating people, or persons are just bundles, in which case person-centered approaches also require a principle of unity for bundles. The positive reductive account is grounded in the causal connections among the neural realizers of the bundles of mental states and in the internal coherence among perceptually-based mental states in such bundles. Each of these is ultimately "based upon the spatio-temporal continuity of the body which is their point of origin" (p. 437). The details of this latter suggestion are not developed in this paper, but begin with the fact that our perceptual states "reflect a point of view, a perspective on the world" by virtue of being tied to a single bodily trajectory (see Bermudez, 1998).

Many papers in *Models* emphasize the importance of the body and its environment in an adequate account of the self. J. Cole argues for the importance of embodiment by exploring disorders in which patients lose the ability to perceive or to execute facial expressions. Cole considers cases of

autism, blindness (congenital and adult), Möbius syndrome, and Bell's Palsy. (The latter two can involve paralysis of facial muscles.) Cole collects patient reports revealing how such disabilities affect one's understanding of emotion, one's sense of integration into a community, and ultimately, one's sense of self. Cole does little to tie his discussion together with traditional philosophical or scientific work on the self, and one begins to wonder exactly how the development of a personality or of a social self is related to, e.g. the sort of self discussed by RH or by Strawson. However, two papers on human development, one by G. Butterworth and the other by M. Legerstee, dovetail nicely with Cole's discussion. Each stresses the importance of the environment in the ontogenic development of the self. Both stress developmental evidence that the mental, reflective self (of the sort emphasized by Strawson) is an ontogenic late-comer. Butterworth argues that this latecomer is "founded on and remains dependent upon" (p. 204) the stability of the visual environment, kinesthetic and proprioceptive awareness of the body, and one's emotional point of view in interpersonal relations. Legerstee describes the results of her eye-gaze experiments with infants, arguing that infants (at six months or earlier) distinguish themselves from their environments, recognize their own voices, imitate facial expressions, recognize themselves in a mirror, and can predict the behavior of others on the basis of a rudimentary theory of mind. She takes this as evidence for "the universal awareness of infants' physical, social and mental selves during their first six months of life" (p. 227).

Finally, Jennifer Radden considers the implications of "pathologically divided minds," including cases of Multiple Personality Disorder/Disassociative Identity Disorder (MPD/DID) and thought insertion, for the synchronic unity of the self. Does Eve have three selves, one for each face? In what sense (if any) were the Son of Sam's murderous thoughts not his own? Radden offers a unified treatment of "co-presence" in MPD/DID (when one personality passively experiences or witnesses the actions and thoughts of other personalities) and of thought insertion in, for example, schizophrenia. Neither case, she argues, should lead us to believe that these patients, at any given moment, are more than one person; Eve's subjective experience of her multiples' thoughts guarantees, by definition of "experience," that the thought is, at least in some sense, hers. The same

goes for Berkowitz's experience of his neighbor's dog's thoughts. Yet, Radden argues (borrowing from the work of Frankfurt, 1976, and Graham & Stephens, 1994), there is a further psychological fact that helps to make sense of these puzzling cases: although Eve experienced the thought, the thought was not a mental action of Eve's. A careful look at pathologically divided minds thus leads us to recognize two dissociable elements of the sense of self: self as the experiencing subject of thoughts and self as the agent of thoughts.

Where does this leave us when the party's over? There are many interesting questions that can be framed with the word "self." What is the nature of the self? What is the relationship between selves, people, and living human organisms? How does the sense of self develop over time? And how might it be degraded through disease or brain damage? What are the criteria of identity for selves over time? Is the self anything over and above the diverse activities of the systems that compose our bodies? How ought one study the self? How could the self have evolved through natural selection? Does the self end at the skin, or does it extend out into our physical and social environments? How do we come to be able to distinguish ourselves from other things? How do we come to recognize ourselves? As interesting as these questions are, they are not all equally related to one another. Without some effort to sort them out, there is a real possibility that when the last song plays, each of the guests to this party will be dancing with a different shadow.

### Acknowledgements

Thank you to the students of my Fall 2001 Philosophy, Neuroscience and Psychology seminar for a semester-length discussion of this book and related articles. I owe a special debt to Sarah Bersen, Philip Meier, Kevin Vallier and Suzanne Wood.

### References

- BERMUDEZ, J.L. (1998). *The paradox of self consciousness*. Cambridge, MA: MIT Press.
- CAMPBELL, J. (1994). *Past, space and self*. Cambridge, MA: MIT Press.
- DENNETT, D.C. (1991). *Consciousness explained*. Boston, MA: Little, Brown.
- FRANKFURT, H. (1976). Identification and externality. In A. RORTY (Ed.) *The Identities of Persons*. Berkeley: University of California Press.
- GRAHAM, G. & STEPHENS, G.L. (1994). Mind and mine. In G. GRAHAM & G.L. STEPHENS (Eds) *Philosophical psychopathology*. Cambridge, MA: MIT Press.
- NEISSER, U. (1994). Five kinds of self knowledge. *Philosophical Psychology*, 1, 35-59.
- CARL F. CRAVER  
*Philosophy-Neuroscience-Psychology*  
 Washington University  
 One Brookings Drive  
 Campus Box 1073  
 St. Louis, MO 63130, USA