

How Can the Self Understand Itself?

A Review of *Models of the Self* Edited by Shaun Gallacher and Jonathan Shear

John G Taylor
Department of Mathematics
King's College
Strand, London WC2R2LS
U.K.

john.g.taylor@kcl.ac.uk

Copyright (c) John G. Taylor 2002

PSYCHE, 8(12), October 2002
<http://psyche.cs.monash.edu.au/v8/psyche-8-12-taylor.html>

KEYWORDS: schizophrenia, inner experience, self, phenomenology, pathologies of self, Eastern meditation, pure consciousness, robots, explanatory gap, agent, consciousness.

REVIEW OF: Shaun Gallagher and Jonathan Shear (Eds). (2000). *Models of Self*. Thorverton, Devon, UK: Imprint Academic. \$39 hbk. 524 pp. ISBN: 0907845096.

ABSTRACT: The self has been debated about ever since the emergence of philosophy in classical antiquity, and more recently has become the subject of intense study in the understanding and possible amelioration of mental diseases such as schizophrenia. An effective model of the self is needed to understand the breakdown of inner experience and its relation to content in such mental diseases, and how this might be reversed. This book is full of important insights into the self which will play a role in future development of such understanding. Its appearance shows that there is progress in our understanding of that subtle component of our experience: it is an important contribution to the literature on self.

1. Introduction

This is an excellent book on a very deep and important subject. The self has been debated about ever since the emergence of philosophy in classical antiquity, and more recently it has become an object of concern in the study and amelioration of mental disease. Schizophrenia and Alzheimer's disease, for example, take their toll on the normally functioning self, so that a disturbed person finds it increasingly difficult to come to terms

with the breakdown of their inner experience and their ability to relate to the social world in which they had previously played an effective role. An effective model of the self is needed to understand how this breakdown of inner experience and relation to other in the world is occurring and how it can be reversed. This book is full of important insights into the self which will play a role in future development of such understanding. Let me describe in a more detailed manner what the contents are before I turn to my assessment of the effectiveness of the models themselves, and the possibility that some sort of consensus might finally be emerging about the self.

The book is a rich meal, divided into six courses comprising over 500 pages: philosophical controversies, cognitive and neuroscientific model, developmental and phenomenological constraints, pathologies of the self, meditation-based approaches, further methodological questions, ending with a reply by Strawson as keynote author. Not all the chapters address the first keynote article, but they all add to the menu by casting useful light onto the nature and analysis of self.

The book commences with a brief but stimulating survey of the problem of self by the editors. They mention, among many features:

- i. Descartes proposal that the self is a 'single, simple, continuing and accessible mental substance',
- ii. Locke's later suggestion that consciousness and the unity of self is provided by memory,
- iii. Hume's inability to discover the self by introspection: 'I never catch myself at any time without a perception',
- iv. The range of modern responses to Hume's denial of the existence of a separate self imply the lack of consensus, and include: claims that the self does not exist, or that it can only be searched for by brain science, that it is purely a grammatical fiction, or it is the centre of social narratives, and other possibilities.

That there is still a range of perspectives on the self is amply borne out by the various chapters in the book. The first, keynote chapter is a personal one by Galen Strawson. He starts by considering eight components of the self, as a single (both at a given time and across time) mental thing, distinct from all other things, it is a subject of experience, an agent, an possessing a personality. This leads to a set of four questions on the nature, especially for a human, of the sense of self, of the grounds for possessing such a sense, and the possibility of such a thing as self. He bases his discussion on the need to found consideration of self on personal experience - on the phenomenology of self. Strawson provides an insightful phenomenology, at the most primitive level: experience is composed of a string of hiatuses or gaps, like a 'string of pearls'. He then relates this view to similar considerations arising from Eastern meditation. Such a primitive experience is more basic than that of oneself as a person; it is the pre-reflective self explored by Western phenomenologists. Thus each of us is composed of a strings of selves, and so are composed of many mental states.

A perceptive development of Strawson's position is that of Katherine Wilkes, who suggests that unity of self over time is needed to agree with our own experience. She bases this argument on patients who continue to be aware of themselves from moment to moment but have to be introduced anew to the medical staff of their hospital each time they meet them. As she points out, there is no 'person' in these patients. A very different point of view is taken by Eric Olson in the next chapter. He claims that, as his title states: 'There is No Problem of the Self', since there is nothing one can point to and say 'The self is one of those'. He suggests that philosophical analyses that claim to be about the self are actually about other things all together, such as personal identity, semantics, moral or cognitive psychology or related topics; they are not about the self per se. The problems they attempt to solve could all be posed without ever using the word 'self'. Another avenue altogether to Strawson's is followed by John Pickering when he claims 'Selfhood arises from the assimilation of cultural signs by a semiotic process that is a fundamental process of nature'. Starting from Strawson's analysis he adds to it by emphasising the need for discussion of process and interdependence, and especially of interaction of the individual with the environment. He thereby underlines the difficulty of 'ascribing consciousness to events within the head alone'.

In the succeeding section there are articles by distinguished neuroscientists. The first, by Ramachandran and Hirstein, proposes three 'laws' of qualia they deduce from various neurological syndromes: qualia are irrevocable, they do not always produce the same behaviour, and thirdly qualia endure in short-term memory. This leads them to propose that qualia evolved in order to facilitate non-automatic, decision-based actions. In the following article Jaan Panskepp proposes that a primitive, affective sense of self arises in certain brain stem circuits. This is suggested to cause primitive emotional states - pleasure, lust, hunger, pain. The relevant neural system is claimed to lie at the periconscious core of all forms of animal consciousness, and so help close the explanatory gap between brain circuits and the psychological nature of affective feelings. Don Perlis takes a very different tack, based more on artificial intelligence. He proposes that 'consciousness is the function or process that allows a system to distinguish itself from the rest of the world'. It is a special kind of self-reference, that involved with a system that can refer to itself as the entity that is doing the referring. He comments that a self-referring machine has never been built, but does not see there to be a barrier in the way of constructing such a machine. Jun Tani is even more machine-oriented, using dynamical systems language to attack the possible nature of self in an intelligent agent. He suggests, from considerable experience in developing navigating robots, that selfhood of such agents arises from the co-existence of stability and instability in goal-directedness. The self, he proposes, becomes aware when incoherence arises in the learning process for interaction with the external world. This embodiment in a robotic system is original and thought-provoking. James Blachowicz carefully considers the nature of inner speech. He considers the proposal that it involves two main partners, one carrying logical articulation, the other experienced meaning. He compares this 'dual partner' view, emphasising equally both partners, to the work of others (Gazzaniga, Dennett, Jaynes) and concludes that 'we can be in two minds quite independent of reference to the minds of others'.

The section on development contains papers by George Butterworth (in which five aspects of self are analysed from a developmental perspective), by Maria Legerstee (again on development of awareness in infants, by Maxine Sheets-Johnstone (returning again to Strawson's chapter and develops a deep phenomenological analysis from a Husserlian point of view), and a perceptive article on self by Zahavi and Parnas on phenomenal consciousness, again seen from the phenomenologists point of view. This latter article gives a valuable perspective on the pre-reflective self, taking it beyond the views of the phenomenologists Merlau-Ponty, Frank and Henry, and raising the deep problem of how to reconcile this pre-reflective self with the content of experience of the external world. The former - the pre-reflective self - is non-relational whilst the intentional world is only relational. This critical gap between inside and outside in the mind forms the basis of the 'hard' problem of explaining consciousness itself: how can there be an intrinsic self but one also able to intentionally move into the world. The apparatus constructed by the authors enables them to pin-point, for example, severe difficulties in the Higher Order Thought approach to awareness. They also offer perceptive comments on the nature of a schizophrenics experience.

Sean Gallagher and Tony Marcel offer a valuable chapter at the commencement of the following section on Pathologies of Self. They are concerned about the need to include context in all analyses of consciousness: they described cases of patients whose deficits of awareness were considerably ameliorated by embedding in suitable action contexts. An example cited was the woman patient who had difficulty lifting and manipulating a cylinder the size of a glass of water but who returned almost to normal manipulative powers when serving mugs of tea to guests in her home. They develop these observations with respect to the intentional attitude, in terms of the attentional focus (on an action, for example, compared to its significance) or the breadth of such focus (inner or outer directed, on the self or the world, is purely on a single action or goes beyond that to its temporal and social significance), and in the mode of attention (immersed in the action or also watching it so non-immersed). They also consider the problem raised by assumed unity of consciousness, quoting earlier work of Tony Marcel's indicating that such a unity can be shown to dissociate under suitably degraded experimental conditions. These remarks on the importance of context being included in discussions of consciousness lead the authors to the important conclusion that during normal day-to-day activity the reflective/reflexive self are 'in many respects non-operative (p 289). They ask how one can capture the pre-reflective self, involved as it is in the humdrum or on-line consciousness very distinct from the retrospective state being probed by psychologists in the lab or considered more loftily by philosophers. They also consider ecological aspects of the self as well as aspects of the ethical self. The following article by Jonathan Cole is very different, considering the effect on their experience of selfhood of subjects with facial recognition difficulties of various sorts. These can have enormous effect, not necessarily always detrimental, as in the case of the student with bilateral Bell's palsy. He discovered that he gained confidence hiding behind his immobile face, and became more forceful and confident. Cole's conclusion of his analysis of these cases is that they can lead to enormous impoverishment of self. The following article by Louis Sass is a perceptive one on schizophrenia and self. He describes vividly the dual nature of the schizophrenics experience: loss of integrated intentionality compared to hyper-self-

consciousness. He considers writings from Foucault and Tausk as well as a schizophrenic patient who played out in his description of his own case the dual aspects of schizophrenia enunciated above. He concludes perceptively with a neurobiological discussion which equates the duality mentioned above with both hypofrontality and hyperactivity in posterior sites in schizophrenics. The final article in the section is by Jennifer Radden, and is on divided minds. She points out that such divided minds can only be explained as involving misguided response to experience on the basis of mind models which regard the unity of consciousness as a logical feature of the world; these are to be compared with those models which consider the unity of the individual as a psychological fact. There results an important research program to determine the true state of affairs in the case of disowned minds, thereby leading to tests of the two opposing theoretical approaches to mind.

The following section turns to an area of self of great interest but full of undoubtedly controversial aspects: meditation. Leading practitioners appear here, and offer intriguing glimpses of their esoteric practices. Robert Forman leads off with a challenging article: 'What does mysticism teach us about consciousness?' He gives a tantalizing glimpse of the depth of the meditative experience in a number of quotations from great mystics from the past, as well as from his own experience. These are encompassed in the well-known states of 'pure consciousness', the 'dual mystical state' and the oceanic mystical state'. These are well described (as well as they can ever be for states with little or no content). His conclusion is that there are mental states which are non-intentional but still conscious, and that content is not all there is to consciousness. Jeremy Hayward continues by describing a particular Buddhist approach to self, that of rDzogs-chen. He points out the similarity between Strawson's view of self as composed of a 'string of pearls' and that arising from Buddhist analysis. He goes beyond this to support the particle-field view of self - the localization of self in the head gives it a particulate aspect but, more deeply, Buddhist analysis provides a field-like character for experience. The approach leads to experience regarded arising from a discontinuous dynamic process of momentary events, with self arising from repetitions of these dynamic patterns. Steven Laycock builds on this dynamic position, and the main aspect that consciousness does not 'appear' to itself: the 'I' looks like nothing at all. After a discussion of various Buddhist writers as well as Sartre, Laycock concludes that consciousness itself is but an adventitious haze, a sheen. Jonathan Shear continues, in the following article, the take on this problem of the nature of the pre-reflective self. He demonstrates by analysis of the Eastern literature that the self cannot see itself but exists as darkness, supporting Laycock's position. He shows how this phenomenological result can resolve major Western philosophical problems. To Shear (as to many others) the self is qualityless pure consciousness; it is also straightforwardly related to Strawson's account of his own experience that he described in his keynote article. The concluding article in the section by Arthur Deikman, succinctly makes the same points as by Hayward, Laycock and Shear: my 'I' is awareness itself, not that which is being observed; he supports this by comparison between Western and mystic writings.

In the final section further methodological questions are considered. These are by Bermudez (on problems in the reductionist program of mind), by Edey (on subject and

object), by Gendler (on exceptional persons and the limits of imaginary cases involving brain transplants, etc), by Mary Midgley (On being scientific about ourselves, with a special diatribe against the explanatory power of memes),and a final chapter by Galen Strawson as an answer to the comments made about his first chapter by others in the book. He strongly backs his earlier claim that selves are physical objects and that there are many short-lived selves. He also discusses features of this self carefully to conclude his stimulating article.

Having taken you through this very rich feast on the self presented in this wonderful book, how can I give my own assessment? How does it help someone who is actively trying to build systems out of neural network which can be said to possess a self? Of course that is not the explicit purpose of the book. Does this book, which is called *Models of the Self*, enable one to actually construct a working model of the self? If not then in principle it could be charged under the Trades Description Act in any self-respecting country. Let me consider if this book is chargeable on those grounds. Several related questions are considered below.

2. Consensus on Self?

Has the original lack of consensus about the nature of self become resolved? Is there now a sense of unity about its nature? There is some progress on this, at least evidenced by the contents in the book. The leading contender, Galen Strawson, agrees in some detail with the Eastern approach ably presented by Forman, Shear, Laycock and Hayward and there is no serious disagreement with most of the other writers about its underpinning nature. This is supported by Strawson's final chapter and his general agreement with other contributors. Such consensus that I sense is an important step forward, related, I suspect, to the growing understanding coming from Eastern meditative practices, of the nature of self gained by deep and inward searching of it as a guided experience.

3. Presence of a Believable 'Model' of Self?

A model is supposed to instantiate a particular concept, and help explicate it. A model of self should be able to indicate in what way self could arise from the interaction of underlying structures. Strawson presents a possible model, but he goes far beyond phenomenology to do so. He introduces a so-called 'U-field', somewhat akin to a boson or other elementary particle, with a related field character; he even introduces quantum field characteristics for the U-field. But this is completely unsupported by any evidence, and gives no hint as to how this U-field can actually be measured. Are there U-field detectors? Can we accelerate a U-field and observe its interactions with other U-fields, as we can with particle beams, say at the CERN particle accelerator? The common answer to all these questions are: No! Nor can one understand why the experience of self should

arise by the interaction of this newly defined Ur field with ordinary matter in the neurons of the brain. Is it in the synaptic connections that his interaction comes into play? Is it in particular sites involved with self as observed through brain imaging (as in the temporal and parietal lobes for episodic memory)? The Ur-field model is not helpful, as far as I can see, in leading to new understanding of self. All it does is restate biases as to the separateness of self from brain.

Other contributors do give hints as to the material basis (if it exists) of the self: the papers of Tani and of Ramachandran and Hirstein, of Panksepp and of Perlis. Ramachandran and Hirstein suggest a control process at work in the creation of qualia, say in the amygdala and anterior cingulate. However the details of this are very sparse. Tani uses a robot navigating to indicate how control may produce a sense of self. However his approach does not seem to me to have any hint as to how the 'what it is like to be' the robot arises at any point in the demonstration. Panksepp does a valuable job in directing attention to brain stem regions for the creation of low-level representations of self. However I can ask the same question of this model: where does the 'what it is like to be' character of experience arise? Even more forcefully are cases of destruction of the thalamic to cortical circuitry. The cortex gradually dies out, leaving brain stem control emerging. The resultant behaviour is one of an automaton: one only able to respond in a reflexive manner, with no recognition, for example, of loved ones. A graphic description of one such case by his son drove home for me the crucial nature of cortex in the creation of conscious non-reflexive behaviour. Perlis, as I reviewed above, makes a good case for self-referent machinery at the base of self, but does not give any hint as to how to model such a characteristic.

Thus my conclusion from this survey is that there is no acceptable model of self presented in the book, in spite of the excellence of the chapters on its nature. They just do not add up to a model able to be used, for example, to make predictions about experience of schizophrenics.

4. What is the Relation to Neuropsychology?

The relation between the discussions of self presented in the book and knowledge arising from modern neuropsychology on the nature of self is relevant in assessing how well the models are grounded in the world of experience. Deficits in self and breakdowns in its experience in association with schizophrenia and autism were described in several chapters in the book, as well as important developmental features in infants. There was also brief discussions of neuroscientific knowledge relevant to self: brain imaging in normals, patient populations, infants associated with various states involved in the self; neuroscientific and psychological ideas associated with working memory and attention related to consciousness and meditative states. However this was skimpy. The gap between the philosophers and the scientists is still large. It may be because of that the 'models' of self presented were not ones able to be used in any scientific exercise..

5. Is the Self Amenable to Scientific Analysis?

The answer to this question is one that deserves more discussion than presented in the book. It involves the important distinction between the ‘soft’ and ‘hard’ problems so clearly expounded by David Chalmers, and related to the explanatory gap. The section on meditative practice and experience was an indication of things to come. But yet how that could occur is not hinted at anywhere. One exception to this is in the article by Perlis on self-referent machines, but this is not linked in any way to the meditative considerations. Nor does it seem clear from the book as to how that could be achieved.

6. Is There any Function for Self?

This is a crucial question that any model should be able to answer. The situation in general is that there are still two camps: those who consider consciousness as a pure epiphenomenon, and those who consider it gives value in terms of some predictive/control manner. The contributions in the book do not help move one towards an answer. In fact the meditative features explored so usefully could only cloud the water even further: if pure consciousness (or similar meditative states) is absence, this seems to have counter-survival value. Consciousness is meant to give added value by enabling one to make decisions involving many items of information in the brain, combined by some form of data fusion. Pure consciousness, at least in its description as ‘absence’, would seem to reverse that ability. The survival value of pure consciousness is not discussed anywhere in the book (nor elsewhere, as far as I know). So how can it have evolved? Is the pure state an artificial extension of ordinary day-to-day consciousness? As I have suggested in two recent papers (2002a, 2002b), if we do not have an answer to this question then it is unclear that we can obtain a valid model for self at any level.

7. Conclusions

I can strongly recommend this book as a very important addition to the literature on self. Its appearance shows there is progress in our understanding of that subtle component of our experience. However this progress has not gone far enough to accept the claim that the discussion is truly of ‘models’ of self. There are exciting but hazy views of self that one can catch from the excellent contributions in the book. I think that is all we can hope for yet: the time for the self to come out into the open and understand itself is still to arrive. In the meantime we should be thankful for the tantalizing glimpses of it in this book.

References

Taylor, J. (2002a). Paying attention to consciousness. *Trends in Cognitive Sciences*, 6, 206-210.

Taylor, J. (2002b). From matter to mind. *Journal of Consciousness Studies*, 9, 3-22.