The Constructed and the Secret Self

In a justly famous article, Bernard Williams (1970) opposes two decisively different descriptions of what seem to be identical situations. The first description is that of the science fictional case of 'swapping selves': we are to imagine a machine that can somehow extract information about the physical bases of memories, character traits, intellectual and creative abilities, etc. from a human brain and then 'transfer' these to another brain (presumably by reconfiguring it to encode the information retrieved from the first brain). Thus the machine could also *exchange* all mental characteristics between two brains. To make the thought experiment vivid, Williams imagines that a pair of persons, A and B, must choose – and they must choose selfishly and before the exchange – between either receiving a million dollars (Williams's example is actually $100,000, but we must allow for inflation) or being severely tortured, where these choices come into effect *after* the operation. But A and B can only choose *who* receives these outcomes by pointing to the present *body* of either A or B. Thus if A thinks that *he* will be transferred to B's present body, he will do well to choose that his own *present* body be the one that will be tortured and that the person who will inhabit B's present body shall be the one to receive the million dollars. Of course, B's situation is the same as A's, but reversed.

The other description is less hopeful. Here, A is told that he shall be tortured tomorrow, but only after a series of preliminary indignities are inflicted upon him. He shall have his memories erased, his character, abilities, etc. effaced, followed by the implantation in his brain of a set of false memories, a false character, etc. Then *he* shall be mercilessly tortured.

The point is that, leaving aside tendentious talk of 'transferring the self', or question begging accounts of *who* shall be tortured, both descriptions can be seen to describe the same set of events. On the other hand, it seems unquestionable that after the operation there is a fact of the matter about who is who. Either A is still in the A-body, or he is in the B-body (or perhaps, if unlikely, he is nowhere, having been destroyed by the operation). Reflection on this thought experiment provides strong grounds for the former disjunct if we antecedently subscribe to some form of materialism by which one's memories, character, abilities, personality traits, etc. all *depend* upon the brain[1]. After all, we allow that memories can fade, be manipulated or altered and that one's character can be radically changed, and we know that such alterations can occur as the result of physical changes in the brain[2]. It seems a good principle that when a range of some

---

[1] The situation is more complex for a dualist. It is possible, I guess, that a non-material self could jump from one body to another (as many have believed in one way or another), and so it is possible that our imagined brain operation could facilitate such jumps (though *why* it should have such a remarkable, extra-physical effect is another question). But notice that we are arguing over the nature of the *substrate* here, so the principle to be enunciated below remains valid for dualist or materialist.

[2] I mention two famous and extreme cases. In 1848 Phinneas Gage had a 3 foot bar of iron pass through his brain as the result of an accidental explosion. Gage survived but his character was severely altered, completely destroying his old life (see Damasio 1994 for this as well as many other examples). A terrible case of induced amnesia resulted from an operation on 'HM' in 1953. The operation, designed to relieve epilepsy (which to some extent it did), involved removal of much of HM's hippocampus and obliterated HM's ability to lay down any new

object's properties depends upon an underlying substrate then the identity of the object goes with the identity of the substrate rather than with the properties themselves. Such a principle is not universally valid; there are certainly entities whose identities consist in their 'structure' no matter what substrate supports them. But we don't think that personal identity is like that. The Toronto Blue Jays could be dis-incorporated, utterly eliminated from the major leagues and then be reconstituted as the *same* team, even though with different players, different coaches and playing in a different stadium (though I guess it has to be in Toronto). That cannot happen to me, though I can be *resurrected*, if the very same relevant substrate is brought back into being.

If we take this view of personal identity and reject dualism then we should identify persons with their bodies (or, better, their brains). And yet, of course, the science fiction cases retain a certain attractiveness. Is this attraction merely an illusion, encouraged by the long years of extreme ignorance about the physical substrate of the self and a variety of influential religious doctrines that exploited it? I think it is a kind of illusion, but one which it is productive to examine. The dual facts that we allow for survival across severe alteration of memory and personality, and yet are also willing to entertain the possibility of identity-transfer suggests that we have an idea of a self somehow intermediate between the self composed of memory and character and the self constituted by the substrate. Such an intermediate self could be shuttled from substrate to substrate, thus allowing for transference of self, but it could also remain the same despite changes in the more superficial personal qualities of memory and character.

I think this idea is less confused than it appears. In fact, a version of it is a natural outgrowth of a representational theory of the mind, which almost demands that we allow for different 'sorts' of selves.

The function of the mind is to help us survive and flourish in the world and it does this in large measure with the help of representations of that world. When we think about what to do, we plot the outcomes of possible actions, whose perceived relevance is gauged relative to our present beliefs about the state of the world, and whose envisaged outcomes depend upon these beliefs as well as beliefs about how the world will change because of these actions. The very possibility of such an activity depends upon our being able to form representations of the way the world is, and the way the world will be after our actions, as well as our being able to appreciate the differences between these representations. It is difficult not to embrace a representational model of the mind which also allows for representations at much less sophisticated cognitive levels than conscious planning. And it is tempting to assert that what marked the emergence of mind and consciousness was exactly the ability to represent the world coupled to the 'ability' to behave in accordance with the representation rather than in more direct response to the world itself.

We tend to think that animals have minds because they do things that lend themselves to a representationalist explanation (or, as philosophers say, explanations based upon the ascription of intentional mental states). Why is the dog barking up that tree? Because it thinks the squirrel is still there (though we have seen the squirrel depart). The dog's representation of the world has diverged from reality, but its behaviour is governed by how it represents the world, not the world.

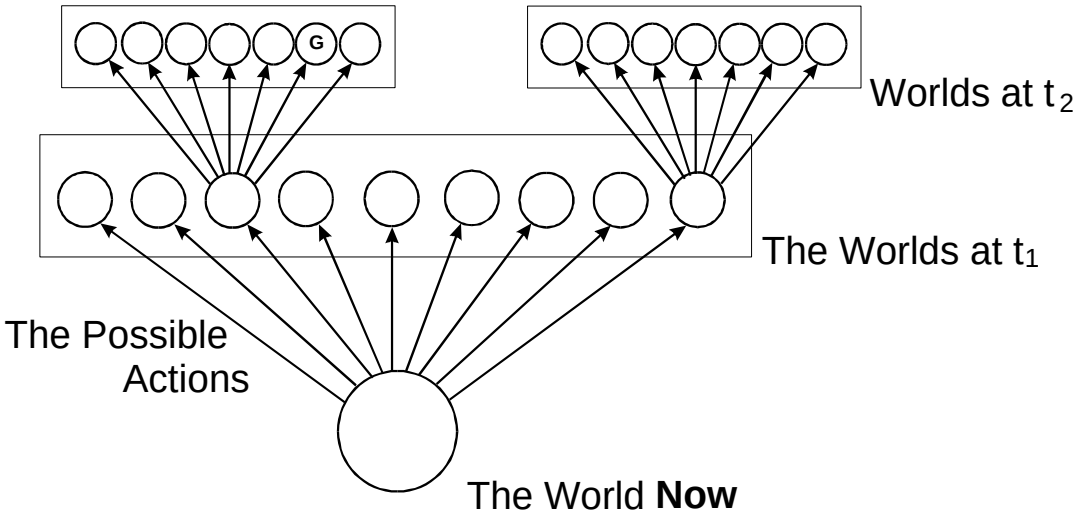
Similarly, machines are more 'intelligent' the more they act according to representations of the world, the less they simply respond (even in complex ways) to the world. AI researchers have explicitly copied our intuitive ideas about the representational mind in their model of

---

memories (see Schacter 1996).

action. Since this model is very simple, it is a worthy metaphor for mind and the problems it highlights turn out to be very deep. Let's outline the model, for every feature of it is ripe with philosophical fruit.

The model is based on the notion of 'generate and search': generate a space of possibilities and search for the ones that are useful. The space of possibilities are the 'worlds' that the system can reach via its actions; usefulness is measured relative to the system's goals. Here is a picture:



(Fig. 1)

Here is the 'game tree' of life. Relative to the current representation of the world, certain actions are possible now, which generate representations of possible worlds whose nature depends upon the actions and the presumed current state of the world. The system can get to the worlds, or at least *thinks* it can get to the worlds, that it can generate by its actions. The action selected is the one that leads to a goal-world (marked with the 'G' in the diagram). You will recognise that this is *exactly* how you play tic-tac-toe.

The diagram tends to imply that the deliberation process is computationally intensive, but the model is more accommodating than appearances suggest. Is it likely that a centre-fielder computes the trajectory of a fly ball from data gathered as the ball is hit, produces a game tree of movements and selects in advance a path through the tree that ends with a world in which the ball his in the glove? Hardly. It has been noticed that when the ball is coming down towards the fielder, he will be where the ball ends up if he moves so that the ball always appears to be coming straight toward him (neither drifting up or down, right or left in the visual field). Performing this feat does not seem to require any computation of Newtonian trajectories. But the process of keeping the ball centred, as it were, in the visual field can be modelled by our picture of deliberation. We need only suppose there is a representation of whether the ball is drifting, and the direction of the drift. The relevant actions are those that reduce the drift, and we have a short loop connecting goal worlds to the representation of the real world.

Nonetheless, this model of thought and action raises very difficult questions. That it has revealed how serious these problems are and also introduced new problems are *virtues* of the model (see Dennett 1984).

For example, any finite system has a necessarily limited representational capacity, which means that its representations are incomplete. Even Borges (1954/1972) famous conceit of the map drawn at 1:1 scale must suffer from severe incompleteness. Incompleteness entails *selection*, immediately raising a host of significant problems about the ways that such selection can be made. In the abstract, these problems are easy to solve: represent those aspects of the world that are *relevant* to the job the representation is supposed to do. In practice, it has turned out to be almost impossible to find a general resolution of these issues. It may be that the constellation of problems of relevance, of which the notorious Frame Problem is one prominent example, are insoluble from within the paradigm of computational cognitive science and artificial intelligence (for an argument to this effect see Horgan and Tienson 1996).

I have no solution to offer but I would like to point out that the 'paradigm-shift' towards connectionism which is often urged in the face of the problems of relevance does not by itself undermine the representational theory of mind. It is true that some champions of connectionism, usually from the group that favours the so-call dynamical systems approach to cognition, have an anti-representationalist agenda (for an introduction to the dynamical systems view of cognition see van Gelder 1995). However, it seems clear that the problems of relevance are problems of representational 'updating' and need not be construed as problems with the representational approach itself, nor even with the general 'game tree' model of cognition. It is only if we regard classical computational mechanisms as the *only* method by which the internal representation can be updated that the problems of relevance arise as a problem for the representational approach. But there is no need to make this assumption. In general, a dynamical systems approach models cognition as a set of interacting variables, perhaps thought of as exerting 'forces' on one another that can typically be described in a set of differential equations. The representation can be regarded as the 'position' of the system in an abstract space and trajectories through this space can be seen as representational updating. It is an exciting possibility – though no more than a possibility – that cognition (or at least certain elements of cognition) can be modelled in this way and that the 'cognitive forces' at work updating the system's representation could solve the problems of relevance, complexity and combinatorial explosion that beset the computational approach. Such a view of cognition would remain within the camp of the representational theory of mind.[3]

Furthermore, it appears evident that the mind does employ representations in cognition and that, short of endorsing a radical eliminativism, any naturalized understanding of the intentional mental states will both use and account for representational states. When you plan your day, for example, you actively construct and compare a variety of representations of the world as it may be at various times in the future, which depend upon how the world is represented to be *now* as well on how you think the world will change depending upon your projected actions. These representations can be false, but they cannot seriously be thought to be non-existent. Much of your consciousness consists in the apprehension of the significance (both in the representational and 'affective' sense) of these representations, along with, though *much* more rarely, an apprehension of the representations themselves. There must be as well a neural implementation of these processes of apprehension, and it is very hard to see how these

---

[3] For an example of the dynamical systems approach to cognition that illustrates how it can be integrated with a representational theory of mind, see the work on 'decision field theory' by Townsend and Busemeyer (1995).

implementations could fail to have representational functions (though how they *get* these functions is another and much vexed question). But anything with a representational function *is* a representation.

Be that as it may, it is the nature of the self from an assumed perspective of a representational theory of mind that I want to focus on here. In particular, I want to explore the way that 'dual' conceptions of the self naturally emerge from a representational approach to the mind. For reasons that will appear, I label the two conceptions of the self the 'secret self' and the 'constructed self'. Let's consider the former first.

If a system is going to successfully manoeuver through the world by use of the game tree model adumbrated above, it will require a representation of itself. Of course, this is not to say that *all* systems that deploy representations in the modulation of their behaviour will perforce contain a self-representation. All sorts of simple control devices – from thermostats to houseflies – regulate their behaviour on the basis of more or less sophisticated representations of the world with no need for a self-representation. But more complex systems that generate representations of the future and various 'paths' (which are sequences of representations) to those represented futures and which must compare their current representation of the world with these paths need to know about themselves: *where* they are, and will be, in the world, *what* properties they are currently possessed of (and how these will change via time and action), what objects *they* can interact with throughout these paths towards the future and so on. Self-knowledge is self-representation. Just as in general the possibility of illusion is like a sign: 'representations at work' so too self-illusion is a sign of self-representation and it is all too evident that cognizers can suffer from a variety of illusions about themselves[4].

There is, however, a 'core' self-representation which is very special. It is needed to account for certain peculiarities of what is called indexical knowledge as well as the relation between such knowledge and non-indexical knowledge, action and perception. The notion of indexical knowledge arises from the appreciation that even exhaustive knowledge of the 'objective facts' will not necessarily lead to any knowledge about oneself *as such*. This can be best explicated by a rather fanciful example (which I borrow from John Perry 1979). Imagine that you are in the ultimate library; every fact about the world can be found within its books. Despite this, if you lack indexical knowledge the library will not reveal to you such mundane facts as who you are, where you are and what time it is. Suppose your name is X. You read in the book before you that X is in the ultimate library. But you, lacking indexical knowledge, don't know that *you* are X. Another book says that X is reading in the ultimate library at 2:30 pm on December 23rd 1998, but since you don't know that you are X, you can't decode this to discover what time it is *right now*. Arguably, unless you already possess *some* indexical knowledge, you cannot deduce *any* such knowledge even from the complete data base of the ultimate library.

Perhaps this seems an implausibly strong claim. After all, as you browse in the ultimate library you might come across a book which reads 'X is reading book Y and X is the only person reading book Y and there is only one copy of Y in the ultimate library (and, by the way, there is only one ultimate library and that is the only place where Y can be found)'. You flip to the cover

---

[4] It is possible that some illusions about the self are actually beneficial. Generally speaking, people think that they are rather nicer, smarter, more attractive and leading more worthwhile lives than they really are, but this illusion might improve their lives. In this case, clear headedness might only increase the suicide rate.

and notice that you are holding book Y, and hence you discover that you are X[5]. However, in order to make this inference you do need one antecedent piece of *indexical* knowledge: I, myself, am holding book Y. Of course, it is not hard, normally, to get this sort of indexical knowledge. It is an interesting question, though, just *how* such indexical knowledge is generated.

I propose that the answer is that there exists a special self-representation (or a self-representational sub-system within the overall cognitive economy) whose function is to transform incoming information into indexical knowledge. This representation is not distinguished by its carrying *any* information about the self. If this self-representation was individuated by the way it represented the self, it would not be the locus of indexical knowledge, for the way it represents the self could then be expressed in non-indexical terms. No list of facts about myself, such as 'Seager lives in Canada', 'Seager has 3 children' etc. could reveal to *me* that *I* am Seager. The self-representation I am positing is primitive and information free. Its distinctive role is to embody indexical knowledge, not by explicitly encoding the information but rather by the way it integrates perception and action with the information already within the system.

Suppose that someone shouts 'Seager's pants are on fire'. I hear this interesting piece of news and my internal representation is appropriately updated, perhaps by the addition of the following item: there is an x, x's name is 'Seager' and x's pants are on fire[6]. But I won't be motivated to do anything about this unless this information is somehow connected to *me*, to *my* concerns. This is the job of the special representation of the self. If, for example, I know that I am Seager, then getting the information that Seager's pants are on fire will lead to the conclusion that *my* pants are on fire. And if I know that my pants are on fire, you can be sure I'll do something about it. (Motivation can arise in a variety of ways but it always depends on the linkage to the self that indexical knowledge provides; in the example just given, if I don't think I am Seager but do think that I am near Seager, I will still be motivated to act – or should be motivated at least – but now my action will be to try to help Seager in some way.)

Logically, we can think of indexical knowledge as working in a perfectly standard way. The posited self-representation can be thought of as a kind of name which functions in inference

---

[5] Note that you will never find a sentence saying 'X is reading *this* book' for that is a piece of indexical knowledge. Maps that can't move (such as the ones in shopping malls) exploit the fact that their immobility renders one item of indexical knowledge unusually secure: you are here (with an 'x' to mark the spot).

[6] I am not assuming that the internal representation is literally composed of sentences in mock predicate logic. I have no idea how our representational machinery is structured; probably it exploits a huge number of distinct modes of representation. I use the sentence format for ease of presentation only. Within this perspective it is natural to regard the self-representation is a kind of name, but one that has a unique function as described above. There are advantages to looking at it this way, notably ease of comprehension and explication. But we are not forced to adopt a 'language of thought' conception of cognition to endorse the idea of a special sort of self-representation. The posited self-representation is distinctive in the way it links together knowledge, perception and action, and this function could be carried out by a wide variety of representational systems. And it seems that in any complex cognitive system such a function must be carred out.

like any other term, so the inference from 'Seager's pants are on fire' to 'my pants are on fire' proceeds by substituting *self* for 'Seager' as licensed by the indexical knowledge that *I* am Seager (as in 'Seager's pants are on fire', '*self* = Seager', therefore, '*self*'s pants are on fire'). It is possible to develop a formal logic in which such self-representations function appropriately as a model of indexical knowledge (see Seager 1990).

What we should focus on here however is the special features of the self-representation which suits it to the job of encoding indexical knowledge. These features are the 'direct' link between the self-representation and both perception and action. If I represent myself as in danger I will act; if I represent Seager being in danger then whether I act or not depends upon whether I know that I am Seager. The self-representation is, so to speak, the ground-floor representation which links me to the world. But I need have no access to this representation and in fact I have no direct access to it. I had to *posit* its existence.

There is something of a paradox here. Insofar as I can consciously represent my self-representation to myself it becomes a bearer of information about myself which may or may not be linked to perception and action. It becomes a name for myself like any other which must itself by linked to the self-representation before I would be motivated to do anything with respect to its conscious representation. This paradox arises because there is nothing to the self-representation *except* its links to action, perception and other knowledge. It can be thought about only by creating a representation of it, which is not identical to it.

Nonetheless, I think the possession of this kind of self-representation is necessary for us to have any sort of fundamental sense of self, for it is what unifies our awareness of the world[7]. Our consciousness of the world is unified in the sense that everything we are aware of we are aware of relative to ourselves, as things that we, ourselves, perceive in one way or another and that we can act 'towards'[8]. It is rather crude to say that so far as I am aware the world is exhaustively catalogued as the things that matter more or less to *me*, but it is not altogether inaccurate. I think this is Wittgenstein's point when he says: 'what brings the self into philosophy is the fact that "the world is my world"' (1921/1961, 5.641). There also seem to be connections between this idea and Kant's transcendental unity of apperception. If we regard Kant's *I think* (which can accompany all other representations) as a gesture towards what I have been calling indexical knowledge the similarity is striking (see Brook 1994, especially ch. 4, for more detailed links between Kant's philosophy and this way of looking at self-representation). A representation is *mine* if it is possible for it to be linked to the self-representation, and anything which I am aware of either via perception or action must meet this condition or be capable of meeting this condition. Consider once more the example of the burning pants. What I can 'notice' about the situation is whatever about it that can forge links to my self-representation. Although I am in fact Seager, I can't be aware that *my* pants are on fire just by knowing that Seager's pants are on fire unless I know that I am Seager. It is always through the link to the self-representation that I

---

[7] Thus I think that simpler creatures that interact with the world without having a self-representation – the sort of creature considered briefly above – cannot develop a sense of self. Of course, much more than possession of the kind of primitive self-representation posited here is needed for a sense of self. What more is required will be discussed below.

[8] Here perception must be taken to include remembering and imagining, and action must include 'thinking about'.

become aware of features of the world around me.

The peculiar 'position' of the self-representation is expressed in another remark of Wittgenstein's: 'the subject does not belong to the world: rather, it is a limit of the world' (1921/1961, 5.633). Thinking of Wittgenstein's 'subject' as our posited self-representation, it is a limit in the sense that nothing unrelated to it can ever enter the world as perceived or conceived. In another sense, it might be thought of a the *centre* of my world, utterly invisible because everything is seen from its vantage point[9].

There are two fundamental assessments of the world that the secret-self must be responsible for: *what is happening*, and *how does it matter to me*. The latter is far more important than the former, for the only things that I really need to know about are the things that might matter to me. It is tempting to speculate that the secret-self trades in speed, so is satisfied with rough assessments of truth and quick judgements of value. Given this speculation it is further tempting to *locate* the secret-self in the 'lower' brain, in the so-called limbic system, perhaps, for one more definite possibility, the perceptual and motor pathways that run through the amygdala, for which there is abundant evidence of rough, quick and decisive 'assessment' of truth and value (see Ledoux 1996, ch. 6 for some fascinating data on fear conditioning and the distinction between the operation of the 'thalamo-amygdala' pathways and the 'cortico-amygdala' pathways). One last speculation would then stress the significance of the relation between the 'old-low' brain and the 'new-high' brain, perhaps going so far as to locate the systems responsible for the *constructed* self (see below) within the latter even as we locate the secret-self sysems within the former.

It can also be pointed out that by their nature there could in fact be *many* secret selves within any subject. These would be distinct centres of awareness, possibly quite disjoint from one another (though nothing prevents the constructed selves – to be discussed below – which will be associated with these secret selves from knowing about the other *constructed* selves). These secret selves would be distinguished primarily by what they 'regard' as true and valuable (see below for the central significance of these notions for both kinds of selves), but they might also differ in the sorts or amount of cognitive resources available to them (e.g. memories, skills, etc.).

This special self-representation is what I called above the secret self, for it is invisible to the subject and we know it only through postulation. Yet of course we know much about ourselves. But this is a different self, which I call the constructed self.

Let us ask how we know about ourselves, our *mental* selves that is (not our bodies), our thoughts and feelings, hopes and dreams as well as the memories that anchor us to our lives. It is correct but facile to reply that we know ourselves by introspection. We then have to wonder how introspection works. The answer to this explains why we should call the self we know by introspection the *constructed* self and reveals the relation between the two selves.

What is in my opinion the best theory of introspection has recently been developed by Fred Dretske (1995, ch. 2). His theory his incomplete however, and I want both to review it and extend it to a more general theory of introspection.

Dretske's idea is that introspection is a form of what he calls 'displaced perception' which is simply learning about one thing by perceiving something else. An example he uses is

---

[9] A better metaphor which rather nicely combines both the idea of centre and limit might be an adaptation of the old mystical view of God. The subject is a sphere whose centre is everywhere and whose circumference is nowhere.

learning that the postman has arrived by perception of the dog's barking. To get such knowledge one must hear the dog and one must also *know* what the dog's barking signifies. Introspective knowledge of our own perceptual states similarly requires that we perceive but also that we know that perceiving is as a mental act. All knowledge – at least all declarative knowledge such as introspection delivers – is conceptual and so requires an appropriate field of concepts for its formulation. In the case of introspective knowledge, what concepts would these be? Since introspective knowledge is knowledge of the mind, they must be mentalistic concepts, concepts of mental states. So introspective knowledge requires the field of concepts that taken together form our notion of the mind. I don't think it does any harm to use the familiar label of *folk psychology* for this body of concepts along with their associated grounds for application. I know that I am perceiving red, when I am perceiving red, because I can apply the concept of *perceiving red* to my perception of red. I don't need to perceive my perceiving (as certain internal scanner theories of introspection have maintained, see for example Armstrong 1968) to make this application any more than I need to perceive my perceiving of a barking dog to apply the concept of 'barking dog' to that object.

Of course, I *do* need to be perceiving red to make the introspective application of the concept 'perceiving red.' In fact, I have to be consciously perceiving, for if I was not conscious of the colour I would have no ground for asserting my introspective knowledge claim (nothing, as it were, to apply my concept to for, as Kant famously said, concepts without intuitions are empty). Without consciousness there is no evidence on which to ground the introspective knowledge claim. We can, I suppose, still imagine bizarre science fiction cases where I come to know that I am, somehow, *unconsciously* perceiving red, but this knowledge would not be introspective knowledge just because there is no conscious mental state to provide the grounds for any introspective knowledge. The point can be made in a partial definition of introspection as self-knowledge of a mental state on the basis of one's state of consciousness engaging one's mentalistic conceptual machinery.

Although my conscious states provide what can be called 'evidence' for my judgements of introspection, it would be misleading to say that I *infer* from my state of consciousness *to* an introspective judgement about that state of consciousness. For this would imply that I already know what my state of consciousness is, which would be to say that I have already introspected. If I infer from anything here, it is from the way the world is presented to me (something I know without introspection). It requires additional conceptual equipment to go from the presentation to the knowledge that the world is being *presented* to me; I need the concept of 'conscious presentation' which is not needed just for the world to *be* presented to me. So, when Dretske talks of 'displaced perception' as a model for introspection we should *not* think of the displacement as involving a move from an awareness of a mental state to a secondary awareness of that state (or yet another mental state), but rather as a move from an awareness of a non-mental state (or object, scene, bodily condition, etc.) to the awareness of the mental state of being aware of that non-mental thing[10].

---

[10] I suppose we could allow that there are some *mental* states that can be perceived in some way independent of introspection. I myself can see little need for this, but if one wanted to assert that feelings (of pain, fear, anger, etc.) were perceptions of mental entities the model could accommodate that desire. These perceptions would not be introspections for they would not be awarenesses of the mental as such (obviously one can feel angry without having any introspective

If we think in terms of inference, we require at least two beliefs: the input to and the output of the inference. If the input of the inference was something like the belief 'I am aware of a tiger in front of me' we would have implicitly appealed to introspective knowledge for we are claiming that I already know about my *awareness* of the tiger. The account offered would thus be circular. We should instead insist that the input belief is 'a tiger is in front of me' and the output belief is 'I am aware of a tiger before me'. Think of children. At an early age they can form the belief that a tiger is in front of them; it takes more conceptual sophistication for them to know that they are *aware* (or are visually aware) of a tiger in front of them. Such increased sophistication is very important for it allows children to entertain the possibility that they might be having nothing more than the mere *visual experience* as of a tiger in front of them and that others might have a divergent experience of the world. The ability to comprehend the epistemic distance between the world and the experience of the world is not some kind of benighted proto-Cartesianism; it is a vital step towards self-consciousness and an awareness of one's own identity.

The key to understanding this position on introspection is always to bear in mind that when we perceive we do not perceive a perceptual state but rather we perceive what the perceptual state represents. Seeing a tiger involves a representation of a tiger but it does not involve *seeing* that representation. Thought is the same; when we think, we are aware – in the first instance at least – of the thought's content, not of the thought itself. To adapt a remark of Dreske's (1995, pp. 100-101), mental representations are the things we are conscious *with*, not the things we are conscious *of*. Although it is venerable, the idea that we are really aware of our mental states instead of being aware of what they represent is as confused as the idea that we can only talk about words because we have to use words whenever we talk. 'Talking about X' involves the use of words but it does not require that we talk *about* those words in order to talk about X. Just so, seeing a tiger demands the use of representations (of tigers) but it does not require that we *see* those representations. The fact that perception can be illusory or hallucinatory is of no more significance than the fact that we can utter falsehoods. Obviously, there is no reason at all to think that the sentence 'tigers live on the moon' is *really* about its own words just because it is false.

The theory is not, then, that one infers from a knowledge of one's state of consciousness to introspective conclusions about that state, though Dretske's examples sometimes unfortunately tend to suggest such an inferential model[11]. The 'evidence' needed for introspective knowledge

---

knowledge that one is angry, or is feeling angry). I think it is preferable to regard sensations as a kind of perception of the *body* which we are aware of as we are aware of the rest of the world (though via distinctive sensory channels). Thus, for example, the sensations of anger involve an awareness of events in the body, as well as strong awareness of *value* and *dis-value* (a vitally important component of consciousness which will be discussed below). An awareness of anger as such requires the level of conceptual sophistication which permits introspection. Even many adults are sometimes in states where they have the feelings of anger but don't know that they are angry (and similarly for other emotions).

[11] This causes Dretske some difficulty when he tries to explain how introspective knowledge is more 'direct' or 'secure' than other sorts of knowledge (see Dretske 1995, pp. 60 ff.). I will discuss this a little more below.

of our own perceptual states is simply our own conscious perception of the *world*, not a consciousness of that consciousness. One can, however, suffer perceptual delusions, illusions and hallucinations. Perceptual consciousness remains throughout all of these, and may be indistinguishable from veridical perception, and so introspective knowledge of our own perceptual states also remains possible, though such knowledge will inherit the illusory aspect. Thus my introspective claim that I am *perceiving* a horse can be in error no less than my perceptual claim that there is a horse in front of me, but I can weaken my introspective claim (e.g., in philosopher's jargon, to something like 'I am in a horse-perceiving-like perceptual state') just as I can weaken my perceptual claim (e.g. to 'there seems to be a horse in front of me'). But my introspective judgement arises in both sorts of case not via conscious inference but rather through the application of the appropriate mentalistic concepts (and, of course, the application of a concept is not itself a conscious act which itself requires some kind of reflective deliberation – such a view leads to an obvious vicious regress)[12]. Once I've got the concept of 'perceiving a horse' I can apply it to the state of perceiving a horse – that is what 'having a concept' *is* – and the 'input' to this application is the conscious perception of the horse (not a perception of the perception of the horse).

On this model, the evident fact that animals and children (at least very young children) are incapable of introspection is neatly explained by their lack of the proper field of mentalistic concepts, the lack of a 'theory' of the mind. They consciously perceive the world, but don't know that that is what they are doing and it is this lack of conceptual machinery that precludes introspection. Their inability to introspect does not stem from a lack of some special internal scanner within their brains (as Armstrong and perhaps Churchland would appear to be forced to aver, see Armstrong 1968, Churchland 1985), nor is it, as Ryle would have it (see 1949, ch. 6), that they cannot perceive their own behaviour nor 'hear' what they say to themselves (children talk to themselves long before they can introspect). To the extent that the conceptual system that makes up our theory of the mind is a relatively late acquisition, to that extent introspective knowledge will itself be a late acquisition. There is evidence that, in all its fullness, the theory of mind is acquired pretty late, typically after age four (see Perner 1991, Gopnik 1993). It is also perhaps worth mentioning that the lack of introspective abilities will not preclude children from making perceptual judgements, for these depend upon a field of concepts which apply to the world, not to the mind itself – and these concepts come first and early.

Note also that if the 'theory of mind' is not only a late developmental event but a late cultural acquisition then we can infer that early hominids were incapable of introspection, and hence were entirely un-self-conscious. The infamous speculations of Julian Jaynes (1976) can be fitted into this model of introspection in interesting ways. To see Jaynes as claiming that it is *self-consciousness* rather than consciousness itself that is a recent, indeed *very* recent, development, and even a recent *invention*, is to make his views rather more plausible than they otherwise appear. It is fascinating to speculate that people began, for example, talking to themselves *before* they knew what they were doing, so they had no way of understanding what was happening to them (Jaynes argues that such people would have assigned such voices to third parties; they were

---

[12] It remains entirely possible that the sub-personal story of concept application will be one of inference-like cognitive processes that marshal evidence and 'test' hypotheses, but the subject is completely unaware of these putative operations. The conceptually informed world is 'delivered' to us without any conscious effort on our part.

the 'voices of the gods')[13].

Dretske's account, as he presents it, is restricted to introspective knowledge of perceptual states but this is only a small province within the realm of self-knowledge. What I want to do now is extend the account to cover other phenomenal states as well as intentional mental states. The case of other phenomenal states requires only a trivial extension. Introspective knowledge of, for example, our own pains requires what we might call first-order consciousness of the pain (which exists in animals no less than ourselves)[14], plus the knowledge that this sort of thing is a pain, or the explicitly second-order knowledge I am in a state that hurts or something along these lines – the exact extent of knowledge of the mind required to underwrite introspective knowledge is of course somewhat vague. Since the phenomenal states provide, by definition, a range of characteristic conscious experience, the displaced consciousness model can straightforwardly apply to them. It is the extension to intentional mental states which is more problematic. It will be achieved by an extension of the realm of the perceptible. What I mean can perhaps best be illustrated by an example.

Suppose I write down a list of simple sums, like, 2+7=9, 12+3=15, 8+5=14, etc. I could ask someone to put a check-mark beside the ones that were *true* and this would be a trivial task so long as my subject knew a little about simple arithmetic. It would *not* be a task demanding introspection. But I could instead ask my subject to check off the sums which he *believed to be true*. This would be no less trivial so long as my subject understood just a little bit about what beliefs were (as well as simple arithmetic). What the subject must minimally understand is an elementary principle of folk psychology, which I can write in a distinctly non-elementary form as: the object of belief qua belief is *the true*. The second of my tasks involves introspection, albeit at a rather primitive level. But seeing that 2+7=9 is correct, or is true, is not in itself an act of introspection any more than is seeing that a zebra is striped. To see what is true, we need to investigate the world, not ourselves[15]. This investigation can occur in the imagination, or via memory, so there is an appropriate internal source for the evidence needed to provide the full range of introspective knowledge of our own belief states. When we 'look' inside ourselves we don't see *beliefs* lined up along our mental hallways, but we can discover *truths* there and if we do discover a truth we have at exactly the same time trapped a belief. Most things are less certain than elementary sums and so we may wonder about our own beliefs insofar as we wonder what is

---

[13] An anecdote about my own daughter perhaps bears this out. When she was three she often complained about not being able to sleep. When we asked her about this she gave the strange explanation that her 'ears were talking to her'. I believe that she was talking to herself but didn't as yet have the notion of such an activity and so could not understand what was going on, and was rather naturally upset by it.

[14] It is important to bear in mind that feeling a pain is not a matter of introspection. Feeling a pain is to be thought of as analogous to seeing a horse (in fact, I believe that feeling a pain *is* a kind of perception, a perception of the body by a highly specialized sensory system). The phrase 'feel a pain' suggests an awareness of a *mental* state, but there is no awareness of a pain as a mental state *as such*; we are just aware of the terrible *feeling*.

[15] Except, pedantry insists, when we ourselves *are* the object of the search for knowledge; this will not always be the search for introspective knowledge.

true, and thus we may be unsure about our own beliefs.

The other basic category of intentional states is desire. The general sort of evidence needed for introspective knowledge here is *value*. The picture needed for the account of introspection I am urging requires that when we look around the world, we not only see objects with their various perceptible properties, but we also perceive a field of values. Is it true that the objects around us come graded in their value to us as they enter consciousness? I think even a little reflection upon experience shows that it is, although this is a multifarious value which is constantly changing in response to all sorts of changes within ourselves. We can use a variant of the arithmetic example to show this. Suppose we replace our list of simple sums with a tray of various small items: some nails, old dry leaves and some sweets and we now ask a hungry subject to pick out the good ones, or the ones that are good to eat. Any subject, over the age of roughly 1.5, could accomplish this. Some (slightly) more sophisticated subjects could be asked to pick out the ones they desire to eat. The first task does not require any introspection, or self-knowledge to be successfully carried out. The second one, as I mean it to be interpreted, is a task involving introspection. (It does require some interpretation since we would normally use the phrase 'pick out the ones you want' to specify the *first* task rather than the second – a significant fact that actually supports the view of introspection I am putting forth[16].) One can consult one's own desires about a field of objects before one selects, but this is simply to gauge the objects' values from the point of view of ascribing desires to oneself, just as taking up a point of view in which one talks of one's own beliefs is to gauge the truthfulness of a variety of propositions (or whatever the abstract object of belief is taken to be).[17]

There are, of course, a myriad of intentional states beyond belief and desire. But it may be that they can mostly be defined in terms of belief and desire, or are just various forms of belief and desire. For example, wishing for p is, more or less, to desire p and to believe that p is unlikely to be (or come) true (see Descartes 1649/1982 for a nice attempt to define a large range of emotions basically in terms of belief and desire). For the really complex interweaving of high level intentional states, the theory I am offering admittedly begins to give out but there a sefl-interpretation theory like Ryle's seems more reasonable. When Mary is trying to decide if she really does love John, she must engage in more than the mere assessment of truth and value. But these two fundamental assessments remain at the core of her self-knowledge. If, for example, she

---

[16] This slight difficulty of interpretation also occurs in perceptual contexts. We often ask for information about the world in explicitly mentalistic terms. For example, if we want to know if a zebra is in the room we can ask: do you *see* a zebra. This is not meant to be a request for an introspective search of our informant's perceptual states, but simply a way of asking if a zebra is visible (it is, of course, by way of such modes of speech that the theory of mind which grounds introspection is passed on to our children).

[17] Moore's 'paradox', that it is somehow senseless for *me* to say 'p is true but that I don't believe it,' even though a lot of other people say this *of* me all the time, is grist for the mill of this account of introspection. On other views of introspection, such as the internal scanner view, Moore's paradox is somewhat troubling for it would seem that I ought to be able to scan my belief states independently of assessing the truth of any proposition and so I could, one might think, quite easily discover that I don't believe something which I can see, so to speak, to be true. The impossibility of this must be given a rather *ad hoc* explanation in an inner scanner theory.

imagines life with John, or goes over the way he acts in a variety of situations, she is assessing truth and value within the imaginary or remembered scenes.

I suggest that there are three classes of 'elementary acts of mind' or consciousness which are required to underwrite this view of introspection. Since we can be in error about the 'external significance' of all three, I will describe them in terms of *seemings*. They are: phenomenal seemings (which encompasses both our conscious perceptual states and all our 'feelings', including pains and other bodily sensations), truth seemings and value seemings. Each one provides a route to introspective knowledge about our own mental states, not via any sort of direct or privileged access but simply by way of the application of mentalistic concepts (drawn from our 'theory' of mental states) to these seemings. Knowing about the mind, I can know that I am in a perceptual state of seeing red when I look at, say, the Canadian flag; knowing about the mind, I can know that I am in pain when I feel the twinge, knowing about the mind, I can know that I believe that 2+2=4 when I understand the truth of this sum; knowing about the mind, I can know that I desire a chocolate when I sense the goodness (relative to the purpose of eating) of the candy before me. To the extent that the three elementary acts are indeed constituents of consciousness we have a reasonable ground for the extension of Dretske's view to the whole of our mental lives. And it does seem to me evident that we are or can be conscious of perceptual properties, truth and value.

Dretske (see 1995 pp. 60 ff.), worries that this view of introspection makes introspective knowledge a species of inferential knowledge, the indirectness of which would to many be objectionably implausible. Dretske seems to believe that the fact that non-veridical perceptions can ground introspective knowledge no less than veridical ones 'neutralizes the objection' (p. 60). He goes on to say that: 'if this is inferential knowledge, it is a strange case of inference: the premises do not have to be true to establish the conclusion' (p. 61). This is a strange way to put the point. Surely the 'premises' here are the seemings I have noted above and, of course, it is *true* that I *seem* to see red even when my perception is non-veridical. We are not really more directly aware of our own mental states than we are aware of the world around us, but within the realm of introspective knowledge we have usually already taken back the epistemic commitment to the veridicality of the perceptual state; at least we are not interested in its veridicality but rather in the perceptual state itself.

The appearance of a direct introspective awareness of our own mental states is to be primarily explained by the fact that the 'inference' from how the world is presented to us to claims about our own mental states is not (or not usually) dependent upon a conscious deliberation but is rather simply the sub-personal application of the mentalistic concepts to their appropriate objects. It is like vision itself – when I see a computer keyboard in front of me I do not 'infer' to the keyboard from an 'appearance as of a keyboard' plus assumptions about the reality of the external world and all the causal relations that link me to it. The concept just springs into my mind and I *see* the keyboard *as* a keyboard. Similarly, when I feel a pain I don't have to think about whether I am *experiencing* a pain; it just springs into my mind that I am.

There are also three features of introspection and consciousness that conspire to enhance the sense of directness in introspection. The first is that we can confuse consciousness with introspection. There is no introspection involved in being conscious of the elementary seemings that make up the 'field' of the mind but, since we, as sophisticated beings in possession of the concept of mind, are aware that we are aware, it can seem that we are directly apprehending our own minds. One can take up a detached view of one's experience and view it *as* experience –

everything then becomes introspectible since one is regarding everything one is aware of as a manifestation of mind. Then the introspective move from world to mind 'disappears' in rather the same way that a constant background noise can disappear from consciousness, simply because it is so ubiquitous. The second feature that makes introspection seem so direct is that we tend to mix together our mental states with the state of the world in our speech; we use the phrase 'I believe …' to report the truth of something (and vice versa); we use the phrase 'I see …' to report that something is before us, and similarly we conflate desire with goodness. We come to see the world *through* our theory of the mind and we are encouraged in this tendency (see note 16 above)[18]. But the third and most important feature is just that the theory of the mind we use to make introspective judgments is second nature to us. We are thoroughly trained in its application from a very early age. We completely and naturally absorb the elementary principles that we *believe* what is true, *want* what is good and are *seeing* what is visible to us.

Into the gap between the thing and the application of a concept to the thing error can creep. A rustling movement in the bush can prompt the firm conviction that a tiger is approaching; here the gap is wide and error is frequent. How can we measure the gap between experiencing a pain and noting that it is a pain that I feel? In one sense (much remarked upon recently) the gap can be as large as we like, for we can imagine that the subject does not properly 'grasp' the concept of pain. This should be uncontroversial. In another sense however the gap can shrink to almost nothing, for introspective knowledge can avail itself of the conceptual resource of demonstrative reference. If I feel a strange sensation, I can always introspect it as a 'something which is happening (to me)'. To entertain the suggestion that I don't understand the concepts involved in such a minimal introspective view is to come very close to undermining thought itself.

Our important self-knowledge is introspective knowledge. It is knowledge about our own beliefs and desires, our pain and joys and most important of all our own memories. To see these as our own, or part of our selves, we must understand what a mind is that it can contain such wonders, and we must learn to recognise them so as to reflect upon them as defining parts of ourselves. I believe that long before we could introspect we felt pain and joy, we could think and plan (though introspection lets us feel more keenly, and sometimes to think and plan more clearly). At some point in the distant past, the concepts which make up the overarching concept of mind must have been invented, as all concepts must be invented[19]. Probably it was not a

---

[18] A word of warning. Philosophers have often confused consciousness with introspection. For example, Locke defines consciousness as 'the perception of what passes in a man's own mind' (1690/1975, bk. 2, ch. 1, p. 115) and more recently Armstrong echoes this with 'consciousness … is simply awareness of our own state of mind' (1968, p. 94). Introspection is entirely distinct from consciousness however. Introspection *depends* upon consciousness but consciousness does not depend upon introspection or the possession of the ability to introspect. I appealed to consciousness in the so-called elementary acts of mind which feed evidence to our introspective abilities, which are grounded in our knowledge of a theory of mind. But obviously I did not explain consciousness itself. And there is no prospect of explaining consciousness by explaining introspection or self-knowledge.

[19] It is perhaps conceivable that some concepts are innate, and have been generated by evolution. But it is not very plausible; concepts have social and normative dimensions that seem

sudden invention but a slow accretion of ideas (there is after all some evidence that certain non-human animals, higher primates in particular, have some kind of idea of self and some sense of their fellows' *mental* states (see Humphrey 1984, Cherney and Seyfarth 1990; for a skeptical account see Heyes 1994). Surely, there was no inventor of folk psychology but there *might* have been (see Sellars's (1956) 'myth of Jones'). The concept of mind was constructed and we learned to apply its many sub-concepts, just as we constructed and learned to apply so many other concepts. The interlocking notions of belief, desire, and feeling are the most precious legacy we have from our remote ancestors, as crucial today as, perhaps, 50,000 years ago (or whenever they were invented – possible *much* more recently).

The self that we know through introspection is a construct. This does *not* follow from the fact that the concept of mind is a construct. *Every* concept is a construct, and a social construct if you like, but that does not make the things to which the concepts apply constructs (still less social constructs). The concept of a 'quasar' is of recent construction but the quasars are not. When we introspect we find mental states, and these are not constructs. But the idea of self has strange powers, for we, being first conscious and second self-conscious beings, can construct ourselves in its image. Quasars cannot do this; they are stuck with being what they are. We, in complete distinction from perhaps everything else in the world, can make ourselves into something for which we had only the image.

What is important to the self? Memories are perhaps the most critical element of selfhood. We treasure our memories, we build systems of external aids to their retention, and we integrate our memories into our own sense of ourselves (we are taught to do so). We also transform our memories in more or less subtle ways. One way we transform them – by selection, distortion, de- and over-emphasis – functions to re-form our own lives (or our images of our lives) into coherent narrative structures (see Dennett 1992, Bruner 1991). Ian Hacking has emphasised that conscious beings have the peculiar power of being able to make themselves into exemplars of some theoretical picture of consciousness or personhood (see Hacking 1995). It is in this strong sense that one's self is a construct, constructed both by oneself continually through life (though usually pretty well set fairly early on) and by the surrounding people (no doubt most notably one's parents) as they exist within a culture that has its own stake in what constitutes a proper self and its own power to bring such selves into being.

But always lurking under the constructed self is the unconstructed secret self, continually monitoring every event for personal significance, ceaselessly linking incoming information to oneself. The secret self is nothing in itself and is vastly influenced by the constructed self (since the construction will influence one's own assessment of value and significance). Such influence is far from complete control however. The assessments of the secret self have some independence, which appears at many levels of cognition. At a fairly low level, in illusions such as the Müller-Lyer, our senses 'tell us' that one line is longer than the other, and though we know that it is not so we still see one line as longer. At much higher levels, we can tell ourselves that we believe a proposition but know that we do not; we can insist to ourselves that something is valueless while knowing that we still desire it. Such conflicts reveal a connection to the secret self, for the values of objects are gauged in relation to myself and my own concerns; they are thus

---

foreign to purely evolutionary processes. *Categories* can be evolutionarily implanted, and can do some of the jobs that concepts do, but nothing as rich as our theory of mind could be reduced to categorization.

linked to the basic self-representation which underlies the consciousness of value. More abstract assessments of value involve much more sophisticated and *self-conscious* appreciation of the structure of the world and my place in it, and this is the home of the constructed self.

I think that we have some sense of ourselves as a 'centre' of perception, truth and value which retains some independence from the self we know by introspection. This centre remains through changes in memory and personality, for without the secret self we are utterly disengaged from the world and could neither act in the world nor even be aware of it. The independence of the secret self gives some currency to the idea of a self separate from the constructed self which could, conceivably, be transferred from body to body without the transfer of the memories, capacities and personality traits characteristic of the constructed self. Such an idea I think helps explain the intuitive appeal of the science fiction cases that began this essay. The idea is however spurious, for the secret self has no *identity* which could persist apart from the bodily mechanisms which underlie the links between perception and action. There is no way a secret self could be transferred from one body to another; everyone's secret self is equivalently *empty* of any traits that could mark it as belonging to one person rather than another. On the other hand, the constructed self could be transferred, but only in the sense of *duplicating* its structures in another.

So in the end we should agree with Williams's final assessment that the identity of persons is the identity of their bodies, while appreciating the ground for the intuition that there is somehow something – a kind of self – *underneath* the constructed self. Paradoxically, the secret self is at once the source of everything that is *mine* in experience, but bears absolutely no relation to anything that makes *me* distinctively *me*. For what makes me, me is the constructed self.

William Seager
University of Toronto at Scarborough

References

Armstrong, David (1968). *A Materialist Theory of the Mind*, London: Routledge and Kegan Paul.

Borge, Jorge (1954/1972). 'On Exactitude in Science', in *A Universal History of Infamy*, N. Giovanni (trans.), New York, Dutton.

Brook, Andrew (1994). *Kant and the Mind*, Cambridge: Cambridge University Press.

Bruner, Jerome (1991). 'The Narrative Construction of Reality', in *Critical Inquiry*, Autumn, pp. 1-21.

Cheney, D. L. and Seyfarth, R. M. 1990 *How Monkeys See the World: Inside the Mind of Another Species*, University of Chicago Press, Chicago.

Churchland, Paul (1985). 'Reduction, Qualia, and the Direct Introspection of Brain States,' *Journal of Philosophy*, 82, pp. 8-28.

Damasio, Antonio (1994). *Descartes' Error: Emotion, Reason, and the Human Brain*. New York: G.P. Putnam.

Dennett, Daniel (1984). 'Cognitive Wheels: The Frame Problem of AI', in C. Hookway (ed.) *Minds, Machines and Evolution*, Cambridge: Cambridge University Press.

Dennett, Daniel (1992). 'The Self as Narrative Center of Gravity', in F. Kessel, P. Cole and D. Johnson (eds.) *Self and Consciousness : Multiple Perspectives*, Hillsdale, NJ: Erlbaum.

Descartes, René (1649/1982). *Passions of the Soul*, translated in E. Haldane and G. Ross (eds.) *The Philosophical Works of Descartes*, Cambridge: Cambridge University Press.

Dretske, Fred (1995). *Naturalizing the Mind*, Cambridge, MA: MIT Press.

Gopnik, Alison (1993). 'How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality,' in *Behavioral and Brain Science*, 16, pp. 1-14. Reprinted in Alvin Goldman (ed.) *Readings in Philosophy and Cognitive Science*, 1993, pp. 315-46, Cambridge, MA: MIT Press.

Hacking, Ian (1995). *Rewriting the Soul*, Princeton: Princeton University Press.

Heyes, C. M. (1994). 'Reflections on Self-recognition in Primates', *Animal-Behaviour*, 47, April, pp. 909-19.

Horgan, Terence and John Tienson (1996). *Connectionism and the Philosophy of Psychology*, Cambridge, MA: MIT Press.

Humphrey, Nicholas (1984). *Consciousness Regained: Chapters in the Development of Mind*, Oxford: Oxford University Press.

Jaynes, Julian (1976). *The Origin of Consciousness in the Breakdown of the Bicameral Mind*, Boston: Houghton Mifflin.

Locke, John (1690/1975). *An Essay Concerning Human Understanding*, London: Basset. My page reference is to the Nidditch edition, Oxford: Oxford University Press.

Perner, Josef (1991). *Understanding the Representational Mind*, Cambridge, MA: MIT Press.

Perry, John (1979). 'The Problem of the Essential Indexical', *Noûs*, 13, pp. 3-21.

Ryle, Gilbert (1949). *The Concept of Mind*, London: Hutchinson & Co.

Seager, William (1990). 'The Logic of Lost Lingens', *Journal of Philosophical Logic*, 19, pp. 407-428.

Sellars, Wilfred (1956). 'Empiricism and the Philosophy of Mind', in H. Feigl and M. Scriven (eds.) *Minnesota Studies in the Philosophy of Science*, v. 1, Minneapolis: University of Minnesota Press. Reprinted in Sellars's *Science, Perception and Reality*, London: Routledge and Kegan Paul, 1963.

Shacter, Daniel (1996). *Searching for Memory: The Brain, the Mind and the Past*, New York: Basic Books.

Townsend, James and Jerome Busemeyer (1995). 'Dynamic Representation of Decision-Making', in R. Port and T. van Gelder (eds.) *Mind as Motion*, Cambridge, MA: MIT Press.

Van Gelder, Timothy (1995). 'What Might Cognition Be, If Not Computation?' *Journal of Philosophy*, 91, 345-381.

Williams, Bernard (1970). 'The Self and the Future', in Williams's *Problems of the Self*, Cambridge: Cambridge University Press.

Wittgenstein, Ludwig (1921/1961). *Tractatus Logico-Philosophicus*, D. Pears and B. McGuiness (trans.), London: Routledge and Kegan Paul.