INTROSPECTION AND THE ELEMENTARY ACTS OF MIND

Once the idea that all mental states are essentially self-intimating is given up, our knowledge of our own mental states requires an explanation, for self-knowledge is undeniably extensive and at least appears to be quite unlike our knowledge of the rest of the world and, most especially, of the mental states of others. The need for this explanation is pressing since all modern theorists of the mind are agreed, for a variety of reasons, in the denial of logically privileged access, incorrigible self-knowledge or self-intimation of mental states, and they allow that many mental states can and do occur without any consciousness of them. This last point naturally raises the issue of the relation between consciousness and introspection, and suggests that there are substantial constraints upon self-knowledge. The problem of introspection, however, is distinct from the problem of consciousness, for a mental state's being conscious is not sufficient for introspective self-knowledge of that mental state, though it does seem to be necessary for such knowledge.

What should we expect from an account or explanation of introspection? Such an account should tell us how we attain introspective knowledge of our own mental states without positing any 'magical' self intimating qualities to them. It should be an account that is consistent with the project of naturalizing the mind, that is, the project of showing how the mind is, or could be, a part of the physical world. An account of introspection should also focus on the fact that introspection is a source of *knowledge*. It should recognize the limits of introspective knowledge, acknowledge the complexity of introspection and carefully distinguish between introspection and consciousness.

Feeling a pain, that is, *consciously* feeling or suffering a pain is not in and by itself a case of introspection; feeling a pain does not by itself constitute any kind of introspective knowledge. Of course pains are conscious states – perhaps necessarily so – but they are not by themselves states of knowledge about the mind. To suppose otherwise would entail crediting all creatures who can feel pain with introspective knowledge about their own minds and while cats, for example, can surely feel pain they do not, I think, engage in introspection. Pains ought to be thought of as analogous to perceptions, as states that provide information about a part of the world, and provide that information in a certain way. Pains provide information about the body. What is interesting about pain is the 'motivational' force which accompanies this form of perception: we might say that pains provide relatively low resolution information with high motivational content, whereas sense perception provides high resolution information with relatively low motivational content. Sense perception is by no means devoid of motivational force however; imagine yourself as very hungry and consider the difference between perceiving a rock and a raisin. As we correctly say, the raisin *looks* good to eat.

Traditionally, sense perception has been seen as the domain of particularly clear *introspective* knowledge. Yet no one should want to say that *seeing a tree* or *listening to the*

wind are cases of introspection or that my knowledge, brought about via sense perception, that, say, the leaves of the tree are green is a kind of introspective knowledge about my own mental state. Introspection can accompany sense perception as a further mental act, permitting us access to our perceptual states of consciousness. The vital importance of consciousness for introspective knowledge is clear here: it seems we cannot introspect states of which we are not conscious. Why not? We *perceive* states of the world that are one and all non-conscious, so why not *introspect* non-conscious states of mind (which are, as noted, now pretty universally allowed to exist and to be possible objects of knowledge)? But I hope this strikes the reader as a very odd question, which should have a trivial answer. And in fact it does, but *not* one that is obvious. A traditional answer *reduces* introspection to consciousness. John Locke, for example, said that '… [c]onsciousness is the perception of what passes in a man's own mind' (1690/1975, bk. 2, ch. 1, p. 115). We have already seen enough to know that such a reduction is unacceptable; Locke's definition might be of *self-consciousness* (and likely that is what he actually intended by the term) but that just is introspection, or at least a form of introspection, so our question remains.

Modern views of introspection divide into three quite distinct forms, which I will label, *perception* theories, *evidence* theories and *transcendental* theories. I think that all of them, and certainly the first two approaches, take their primary task to be the integration of introspection into a naturalistic view of the world, where the main force of the term 'natural' is *anti-Cartesian*, specifically *anti* such Cartesian themes as incorrigible and inevitable self-intimation of our mental states as well as the postulation (if this is the right word) of a class of special mental entities to serve as the objects of introspection, conceived of as a peculiar but sense-like faculty.

There is no space here for a thorough examination of these positions, but I do want to provide brief outlines of them. I will try to point out what I take to be the central *prima facie* difficulties for these theories in order to motivate the theory of introspection to be introduced below, which quite naturally evades these objections.

The most obvious approach to introspection is simply to equate it with perception. According to naturalism, mental states are states of the brain (more or less) and so introspection is perception of the brain. Of course, if a surgeon holds a mirror up so that I can see my brain as she operates on it I am not, in virtue of this unusual perception of my own brain, also introspecting. Just as in the case of perception, introspection requires its own (perhaps only functionally defined) sense organ, which must itself be a part of the brain with its own specialized 'inner receptors.' Thus is born the inner-scanner theory of introspection (see David Armstrong's seminal writings on modern materialism, especially Armstrong 1968, and more recently Paul Churchland, 1981, 1995), which postulates a functional brain system – let's call it the *I-scanner* – designed, by evolution presumably, to actively seek out information about those brain systems which realize lower order mental states. Armstrong explains the action of the I-scanner thus: 'In perception the

brain scans the environment. In awareness of the perception another process in the brain scans that scanning ...' (1968, p. 94)[1]. The I-scanner is by its nature rather more intimately connected to the objects of its perceptual capabilities than are our various sensory 'scanners', and this is supposed to explain both the accuracy of introspection and the long standing philosophical illusion of introspective infallibility. According to I-scanner theorists, introspection is *practically* infallible, but only in the way that perception of near-by objects directly before us in good light, etc. is characteristically infallible. However, as in the case of sense perception, one can envisage at least the philosophical possibility of radical errors in introspection. For example, a science fictional brain surgeon could, in principle, activate a variety of I-scanner states which would yield introspective belief in all sorts of mental states, none of which would actually be present.

One could raise several objections against the I-scanner theory (see Lyons 1986) but the most salient difficulty stems from a curiosity of the phenomenology of introspection – there isn't any (as many philosophers have noted; see for example Lyons 1986, Dretske 1995, Searle 1992). If introspection were properly understood as a kind of perception, one would expect there to be a distinctive phenomenology of introspection; mental states ought to have 'introspectible qualities' just as the objects of perception have perceptible qualities such as colour and form. Unfortunately, it is a plain fact that there is no distinctive introspective phenomenology whatsoever. That you have current introspective access to your perceptual states of consciousness no one will deny, but all the phenomenology you can find is exhausted by the perceptible qualities of the objects of your current perception[2]. Attending to your perceptions may well heighten or otherwise alter your perceptual state but it does not introduce a new realm of introspectible qualities. It would be desperate folly to claim that the introspectible qualities of mental states are an exact *copy* of the perceptible qualities of objects and even worse to claim that we are never aware of anything but introspectible qualities of mental. Lacking the nerve to make this latter assertion, we might ask the I-scanner theorist why, when we introspect our perceptual states of consciousness, the evident 'outerness' of perception is not replaced by an 'innerness' appropriate for introspective awareness of the brain (or, even, the mind)?

The introspectively rather more significant realm of the intentional mental states is still less kind to the I-scanner hypothesis. It is very difficult to convince oneself that beliefs and desires are known by a process analogous to perception. Intentional states do not have a set of quasi-perceptual qualities by which they are internally spotted and distinguished one from another. This is evident in common wisdom. If Mary wonders whether she really loves Tom, she *thinks* about Tom and their relationship. The advice to forget about thinking and just look inside oneself gets one nowhere (Ryle is famously good on this (see 1949; also see Searle 1992, pp. 143 ff.)). Of course, there are various feelings Mary might have when she thinks about Tom and she can be conscious of these as feelings, but the question is, are they *signs* of love. And there, the problem

is that there is nothing more to 'look at'; our intentional mental states do not form an inner world of objects with their various introspectible properties.[3]

But couldn't Mary imagine, for example, what it would be like to live with Tom? She might imagine various domestic scenes or possible situations and 'play out' how Tom might act in these situations. Of course she could and this is no doubt part of what thinking about our own intentional states involves, but this is not introspection. Here again the 'crucial fact' discussed above must be noticed. Consciously imagining something is *not* in and of itself introspection. I can vividly imagine the Canadian flag, with its bright red stripes and maple leaf, but I am not introspecting when I do this. Knowing that and what I am imagining as well as how I am imagining it *is* introspective knowledge, but I do not get this knowledge by somehow perceiving my imagining of the flag. My imagining does not 'look' like anything, though my image of the flag does (at least there is a distinctive, and obviously perceptual, phenomenology of imagining), but 'looking' at my image is just the imagining itself, not an act of introspection, although of course I can, even in imagination, attend to particular aspects of the imagined sensory qualities.

What I shall call 'evidence theories' provide another approach to introspection. The basic claim of evidence theories is that our knowledge of our own mental states is formed in precisely the same way as our knowledge of others' mental states; there is no asymmetry between 'first-person' and 'third-person' knowledge of the mental. Gilbert Ryle was the most forceful exponent of such theories, but Dennett (in some of his guises, see 1987, ch. 4) and Lyons (1986) are also subscribers. Ryle put forth his view succinctly as follows: 'Our knowledge of other people and ourselves depends upon our noticing how they and we behave' (Ryle (1949, p. 181)).

There are some advantages to the evidence theory. It dissolves the surprisingly vexing problem of other minds, or at least transforms it into no less a problem about *other* minds than about our own[4]. It exhibits a nice theoretical economy insofar as since it is given that we do have methods for attributing mental states to others it is rather elegant to enlist these methods for self attribution, thus eliminating any theoretical need for a special and suspect introspective faculty or sense, or even any special mode of self-interpretation. It also retains an acceptable form of 'privileged' access to our own mental states in its allowance that we all generally have access to *more*, if not a different kind of, evidence about ourselves than others, while possessing abundant means to account for the evident failures of self-knowledge.

Now, it is hardly surprising that in some cases people will attribute mental states to themselves on the same basis that they attribute them to others, for people are capable and fond of *thinking about* themselves and their motivations. It does no harm to label the kind of systematic knowledge which we employ in such thinking 'folk psychology' and we quite naturally apply this knowledge to ourselves no less than to others. But this is a frail basis upon which to build an evidence theory of introspection for it simply ignores the vast range of self-knowledge to which

the model applies very badly if at all.

It is, as philosophers have long remarked, extremely easy to gain introspective knowledge of one's own perceptual states but this does not require one to observe, imagine or remember one's own behaviour or utterances. I don't need to say to myself 'I am seeing red' to know that I am in a state of *seeing red*, nor do I need to imagine red (in fact, when one is actually seeing red, it is next to impossible to *imagine* red at the same time, and obviously such imaginings would gain nothing over the perception). Of course, *seeing red* is not by itself an introspective act so perception is not itself an objection to the evidence theories, but introspective knowledge of perceptual states is freely available simply in virtue of the fact that these are conscious states. We simply do not have to watch what we do or say to know that, and what, we are seeing or hearing.

The evidence theory no more plausibly explains introspective knowledge of the vast majority of intentional states than it does our knowledge of perceptual states. Knowing that I believe that, for example, 2 + 2 = 4 is (a rather trivial) case of introspective knowledge. It is ludicrous to suppose that my knowledge of this belief depends upon or stems from an observation of my own behaviour or from some act of imagination (what would it be: imagining myself writing down '2 + 2 =' and then noting that my imaginary self completes the equation with '4'?). It is no different in the case of desire. I know that I like eggplant curry, but not because I find myself ordering it in a restaurant or because I can remember that I have ordered it often in the past or because it is easy to imagine myself ordering it in the future, or because I remember hearing myself say frequently enough 'I like eggplant curry' or because I can easily imagine hearing myself say this, or yet because I frequently say to myself 'I like eggplant curry' (which in fact I do not).

The evidence theory is plausible only for cases of 'difficult' attribution of complex intentional states and only a tiny fraction of my self-knowledge, though I admit a highly significant fraction, involves these. This zone of accuracy for the evidence theory is important. It points to the very real practical problem we face in knowing ourselves as complicated intentional beings within the complex social environment we have constructed for ourselves. A good part of our self-knowledge stems from self-interpretation as the evidence theory claims. But despite the insights it can generate, the evidence theory seems to be inadequate as a general account of introspective knowledge, and is seriously misleading about the fundamental source of introspective knowledge.

In philosophy, it is always at least thought to be possible that the appearance of a problem is *mere* appearance, and so it is with the problem of introspection. The *transcendental* theories of introspection are really more in the nature of a denial that there is a problem of introspective knowledge than a substantive theory of introspection. Of course, such theories admit the existence of extensive self-knowledge and will go so far as to allow that this sort of knowledge has a very

special nature and perhaps has a more secure status than most other kinds of knowledge. But the story of introspection as told by the transcendentalists is quite different from those we have heard so far. The basic idea is that accurate self-knowledge is a condition for the possibility of all attributions of mental states to any subject (hence my adoption of the Kantian use of 'transcendental'). Donald Davidson (1984, 1987) notes that a general failure of self-knowledge would render the interpretation of a subject as believing and desiring incoherent. For suppose that in general some person acted as if they believed, say, that elephants have tusks, even going so far as affirming that elephants have tusks (i.e. uttering 'elephants have tusks') but denied that they believed this or even claimed that they could not tell whether or not they believed that elephants have tusks. Such 'meta-behaviour' would render the initial interpretation of the subject's belief unsustainable and if such 'paradoxes' were the norm, it would certainly throw into doubt the idea that the person had beliefs and desires at all. Note that this sort of *a priori* defence of self-knowledge also leaves room for errors of introspection, so long as these errors remain infrequent and do not become, so to speak, flagrant.

A very elegant and compressed argument for a transcendental theory of introspective knowledge is provided by Akeel Bilgrami at the end of his *Belief and Meaning* (1992, pp. 250 ff.). Bilgrami ingeniously links self-knowledge with the possibility of interpreting people as *moral* agents. Roughly, the proposal is that only those who are aware of, or know, what they are doing can be held morally responsible for the consequences of their actions. Then, Bilgrami argues, since it is a condition for the possibility of taking others to be persons that they be thought of as morally responsible for their actions (at least most of the time), it is a condition for the possibility of taking others to be persons that they (generally) know what they are doing, and one cannot know what one is doing without some insight into the intentional states which actually explain one's actions. Self-knowledge of one's own intentional states thus emerges as a condition of personhood.

There is no doubt that there is something correct in analyses like these, but it is very far from clear that they either solve or dissolve the problem of introspective knowledge. This is because of two related failings which reveal that the transcendental theories cannot evade the requirement to provide a *positive* account of introspection.

For, although the transcendentalist may well be right that generally accurate self-knowledge is a condition of the possibility of ascribing mental states to others[5] and also ourselves, one must nonetheless ask *how* creatures for which this possibility arises are 'structured' so as to realize the possibility. The original transcendentalist, Kant, consigned this structure to an unknowable noumenal realm into which human thought dare not venture on pain of unintelligibility. This is hardly a strategy the modern transcendentalist about self-knowledge should wish to emulate. Nor is it defensible to refuse the request for an explanation of how self-knowledge is

attained simply by appeal to the transcendental proof that, after all, it *must* be attained. Here, the transcendental methodology is closely akin to the use of the anthropic principle in cosmology, which well illustrates the general limitations of a transcendental or anthropic account of some fact. For example, we observe that the Earth is neither too near nor too far from the Sun to support life. This is no surprise, for the Earth being at such a distance is a condition of the possibility of observers. The anthropic, 'transcendental' account of the Earth-Sun distance is, however, no substitute for an explanatory tale of the genesis of the solar system in which each planet attains its appointed place via natural forces, with no help from the fact that the Earth would someday spawn astronomers who can measure the Earth-Sun distance. Every realized possibility must be realized via some 'mechanism'; so our introspective capacities must have some actual source which, when articulated, will amount to a positive theory of introspection.[6]

Furthermore, it is evident that self-knowledge is a cognitive achievement that takes some effort. Self-knowledge, like any other knowledge, does require evidential warrant. When we have such knowledge, we generally also know *why* we know what we know about ourselves[7]. Self-knowledge cannot be freely generated simply by asserting something about oneself and it must answer to the same canons of evidence, warrant and the possibility of error to which any potential domain of knowledge must answer. Further, my self-knowledge must be integrated with my knowledge of others' mental states and my knowledge of the world. But then the range and type of evidence required for self-knowledge is problematic and demands an account.

Although the foregoing discussion is too sparse to be decisive, there are obvious problems with extant accounts of introspection which it would be better to avoid if possible. I believe there is a positive theory of self-knowledge which avoids the difficulties we have canvassed and which also possesses several additional virtues. The account I will advocate will not entirely repudiate either the I-scanner or evidence theories, for both of these grasp some measure of the truth. That is, I imagine that there very probably are brain systems whose function is to monitor other brain systems and it may be that such systems are crucial for the generation of conscious experience. And I am sure that it is also true that we do sometimes gain introspective knowledge by way of the kind of behavioural clues and self-interpretation appealed to by the evidence theories. Furthermore, the proposed view of introspection will accept the transcendentalist's claim that self-knowledge is in some way a condition of (full) personhood; it aims to underwrite the possibility condition rightly recognized in the transcendental theories. The view of introspection I want to defend is an extension of a view sketched by Fred Dretske in his recent Nicod lectures (1995, ch. 2)[8].

Dretske's idea is that introspection is a form of what he calls 'displaced perception' (from which I draw the more general characterization of 'displaced consciousness') which is simply learning about one thing by perceiving something else. An example Dretske uses is 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, so to speak, what perceiving is. Knowledge is conceptual and so requires an appropriate field of concepts for its formulation. Introspective knowledge requires the field of concepts that together form our notion of the mind. As I said above, I don't think it does any harm to label this body of concepts, with their associated grounds for application, folk psychology. I know that I am perceiving red, when I am perceiving red, because I can apply the concept of perceiving red to this instance of my perceptual experience. I don't need to perceive my perceiving (as the I-scanner theory asserts at bottom) to make this application any more than I need to perceive my perceiving of a barking to 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. This answers a question I posed long ago: why is it that we can only introspect *conscious* states of mind (given that there are plenty of mental states that are not conscious)? The simple answer is that without consciousness there is no evidence on which to ground the introspective knowledge claim. We can 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 (this is the trivial, but non-obvious, answer promised above).

So, isn't this just the evidence theory? No, though it is of course *an* evidence theory of introspection, it denies that the evidence needed for introspection is the evidence of behaviour and/or utterance (either public utterance or private talkings to oneself). The evidence needed for introspective knowledge of our own perceptual states is much more direct and traditionally appropriate: our own conscious perception of the world.

Thus, 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[9].

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 (or that there's one in a book); 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* (or otherwise experiencing) 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 Dretske'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* (or be otherwise aware of) 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[10]. The 'evidence' needed for introspective knowledge of our own perceptual states is simply our own conscious perception of the *world*, not a consciousness of that consciousness. One can suffer perceptual delusions, illusions and hallucinations. Perceptual consciousness remains throughout all of these 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').

Dretske's account, as presented, is restricted to introspective knowledge of perceptual states but this is obviously only a small province within the realm of self-knowledge. What I want to do now is to 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 consciousness of the pain, plus the knowledge that this is a pain, or that 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 what is, in effect, 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 *believed to be true*. This would be no less trivial so long as my subject understood 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[11]. 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 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 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? I think 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[12].) 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).[13]

The values at issue here are in a way 'subjective' – they are not properties which others can be faulted for failing to notice (though there is a wide measure of agreement amongst us about what is valuable in any given context). Does this threaten the account of introspection given here? It might if we thought that the subjectivity of the values undermined our attempt to model the awareness of value on perception, on the basis of the claim that such values were properties of the mind rather than the world. But there is no such threat. It is the point of the representational theory of consciousness to distinguish those properties which we represent objects to have by way of an internal representational state from the properties of that internal state itself. The perception of value does not have to be veridical in order for us to claim that the value is represented as being a property of objects in the world (although, I suppose, most such assignments of value are unobjectionable in their context – raisins are good to eat). By analogy, consider the hallucinations of someone suffering acute schizophrenia. These are 'subjective' in the sense that they are internally generated and utterly non-veridical, but they are experienced as being fully 'in the world'. The schizophrenic is in a state that represents the world as being in a certain – highly

bizarre – way; the property being represented is not a property of the sufferer. It may be similar for our perception of value. They are perhaps 'hallucinations', in that there is no objective correlate in the objects around us underlying the represented value-state. These properties would nonetheless remain properties of the world as represented. An exactly similar story can, and has been, told about colour but whether or not the objectivists or subjectivists are right about colour, we still represent colour as a property of 'external' perceptible objects.

One important analogy between sensory perception and the extended perception of truth and value I want to posit is the way first-order error is compatible with second-order introspective correctness. In the case of perception, the main and traditional source of sceptical worries is that, in principle, there is no purely internal way to distinguish hallucination from veridical perception. This is an undeniable, if philosophically unfortunate, fact. But it is an ill wind that blows nobody any good. The fact that our state of perceptual consciousness can be the same in both veridical and hallucinatory experiences means that the application of the mentalistic concepts which constitutes introspective knowledge is the same, and equally justified, in both cases. Thus it is that we can introspectively *know* that we are having a visual experience of a tiger even if there is no tiger present (and even if we *know* there is no tiger present). Similarly, it is possible to believe the false, but in such cases the *false* appears to be the *true*. Since introspective knowledge of one's own beliefs depends in the first instance upon taking something to be true and since such a taking to be true occurs in the case of believing the false no less than believing the true, one's introspective knowledge of one's own beliefs is equally secure in both cases[14]. In the case of desire, much the same story can be told. Whether something *really* has value or not cannot be discerned merely from the appearance of value, but that I desire something can be introspectively known just from this appearance.

Of course, more sophisticated thinkers (and perceivers) know that they can make mistakes, can see what is not there, believe what is false and desire what has no value. They also can be aware of the complexity involved in the attributions of both truth and value. Thus there is room for self-doubt and the machinery of self-interpretation which extends the basic workings of introspective knowledge.

There is 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). And, for the really complex interweaving of high level intentional states, the evidence theory does begin to come into its own. 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

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. It might be worth pausing to consider how such assessments work. Imagining or remembering scenes is to picture the world (not the mind). I think it is as impossible to divorce perceived value from such picturings as it is to visually imagine an object without imagining that it has any shape (or other visually perceptible qualities). Thus when Mary imagines doing something with John (just what is best left to the imagination), she will 'see with her mind's eye' what is going on and also sense the desirability, or otherwise, of the event. This will help her learn something about her own mind. Imagination is also schooled by plausibility, especially if it is a serious exercise in figuring out what might really happen. So the assessment of truth (and its cousin, likelihood) might prevent Mary from seriously entertaining certain imaginary possibilities, and this might tell her something about Tom, her relationship with Tom, and hence may reveal to her something of her own mind. But these matters are very complex, and subject to error. It might turn out that Mary was quite wrong in her imaginative construction of how she would react in certain situations, and discovering what her actual behaviour is could force a bout of Rylean self-interpretation.

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 a '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.

It is perhaps a problem, noted by Dretske (see 1995 pp. 60 ff.), with this view of introspection that it makes introspective knowledge a species of inferential knowledge. 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 'automatic' 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 (given, of course, that my mind has been prepared with the inculcation of the appropriate field of mentalistic concepts).

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 are aware that we are aware, it can seem to a sophisticated consciousness 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 as a manifestation of mind. The introspective inference 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 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. Third, 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 11 above).

To illustrate the operation of the theory and to contrast it with the evidence theory, reconsider the example of my introspective knowledge that I like eggplant curry. One weakness of the example must be dismissed at once. We all carry an extensive store of information about our own minds, and *any* theory of introspection can allow that for a large number of questions about

our own mental states we need do no more than consult this memory store. In particular, on either the displaced consciousness or the evidence theory of introspection, the most straightforward way for me to 'introspect' my liking for eggplant curry is simply for me to remember that I have this liking. So let us assume that somehow the case before us is not one where a memory of an introspected mental state provides the source of (renewed) introspective knowledge. In that case, the evidence theory should assert that I achieve introspective access to my liking for eggplant curry through some kind of observation of my behaviour coupled with an mentalistic interpretation of that behaviour; as noted above, this is highly implausible. The displaced consciousness theory of introspection, on the other hand, depends upon my perceiving the valued qualities (in this case valuable gustatory qualities) of eggplant curry. While the most direct way to get acquaintance with these qualities is to *taste* the curry, I can use my imagination and perceptual memory to experience the curry *in absentia* as it were. If I remember the taste of eggplant curry, I can experience whatever it is about that taste that I value. It is this experience that licences my introspective knowledge that I *like* such tastes. (Here, I want to stress yet again that the remembering of the taste of eggplant curry is not by itself an introspective act; it brings me into contact with a feature (or possible feature) of the curry, not a feature of my mind.[15])

If the nature and the strengths of the theory are now reasonably clear, let me conclude with a discussion of some 'applications' of the theory to various issues in the philosophy of mind, and a warning about the scope of the theory.

The displaced consciousness theory of introspection clearly entails that animals and children (at least very young children) are incapable of introspection because they lack 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 knowledge that precludes introspection. It is not because they lack some special I-scanner within their brains, nor is 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 (and there is evidence that, in all its fullness, it is acquired pretty late, see Perner 1991, Gopnik 1993), to that extent introspective knowledge will itself be a late acquisition. 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.

Here is a more speculative application of the theory. It has been suggested that autism arises fundamentally from an inability to grasp the basics of our commonsense theory of mind (see Baron-Cohen et. al. 1985). It is not known what sort of organic problem could lead to such a specific *conceptual* inability. Nonetheless, there is some evidence that autistic children are incapable (or at least very much less capable than normal children) both of appropriately ascribing

mental states to others on the basis of their behaviour and of deploying mentalistic descriptions in the explanation of behaviour. The displaced consciousness theory of introspection would further predict that autistic children would have severe, and parallel, difficulty in acquiring and using introspective knowledge. But it would also predict that the autistic child could make some progress in the acquisition of a notion of mind, since the basics of introspection are the awareness of the perceptible qualities of the world along with an awareness of truth and value, and autism does not destroy these even if it does alter them in various ways (especially with regard to the emotions). In light of this, it is interesting that autistic children sometimes do find a way to integrate themselves quite successfully into the normal, pervasively mentalistic, world of human society (see Oliver Sacks's account of Temple Grandin in Sacks 1995, ch. 7, which gives some support to the idea that this success is achieved by a much more labourious, intellectualized and still only partial acquisition of commonsense folk psychology)[16].

Currently, there is a debate amongst philosophers, psychologists and cognitive scientists about which of the so-called theory-theory or the simulation theory gives a better account of our knowledge of the mental states of other people (see Carruthers and Smith 1996 for a variety of views on this issue). The theory-theory asserts that we ascribe mental states to others by the application of a theory of mind (a folk theory that is) to their observed behaviour and utterances. So long as we don't put too much weight on the term 'theory', a lot of philosophers would seem to line up on the side of the theory-theory. Obviously, the Rylean view is a version of the theory-theory insofar as it admits and requires the existence of an articulable set of principles by which we attribute mental states to others (and ourselves for that matter). The distinct versions of interpretation theory put forth by various philosophers (e.g. Davidson and Dennett) are also versions of the theory-theory, in which, roughly speaking, the primary principles of the theory are those that maximize the rationality of our fellows' behaviour and mental life. Opposed to the theory-theory, the simulation theory asserts that we ascribe mental states to others by a process of internally modelling the situations other people find themselves in, 'reading off' the mental states *we* would be in if we were in those situations and ascribing these states to others.

This is not the place to attempt a survey of this debate. But I will note that the displaced consciousness view of introspection offered here could usefully add to the resources of the theory-theory, which seems to be somewhat mired in a basically, though more 'scientized,' Rylean picture of self-attribution (see for example Gopnik 1993). More important, if the displaced consciousness view of introspection is correct then the status of the simulation theory becomes rather complex and somewhat precarious. The simulation theorists rely upon the fact that we know what our own mental states would be if we were in the simulated situations we take others to actually be in. But I am urging that we know our own mental states because we know a theory of the mind. If so, the simulation theory presupposes the use of a theory of the mind after all.

However, the simulation theory could still be correct if in our attributions of mental states to others we needed to go through the self attribution process in an imaginary situation. This is not particularly implausible. In fact, since in the view of introspection presented here, self attribution does not stem from self observation of our own behaviour or utterances but rather stems from the elementary acts of the conscious mind, putting ourselves in an imaginary situation could give us a new set of such elementary acts to use in our simulated self attribution. The situation is murky, since, after all, even Ryle could allow that we imagine how we would behave in imagined situations as we try to figure out the mental states of those who are really in such situations. The point to stress here is simply that, contrary to the professions of the simulationists, there is no escape from the need for a theory of mind in the simulationists' view, since their reliance on self-knowledge in fact presupposes just such a theory[17].

Finally, the warning. Philosophers have often confused consciousness with introspection, as the quote from Locke in note 1 illustrates particularly vividly. 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. Nonetheless, it is encouraging that introspection does not deepen the mystery of consciousness but can be given an straightforward and plausible account.

William Seager
University of Toronto at Scarborough

Notes

1. I can't help noting that Armstrong continues the above passage with 'Consciousness … is simply awareness of our own state of mind,' thus reiterating an unfortunate view of consciousness which has been distressingly prominent in philosophical reflection. An early example of this tendency can be found in John Locke, who states that '… [c]onsciousness is the perception of what passes in a man's own mind' (1690/1975, bk. 2, ch. 1, p. 115). Obviously, this makes it impossible to be conscious of anything other than states of mind, which is really very implausible. Locke's (and Armstrong's) notion would be better thought of as a definition of self-consciousness or even introspection itself.

2. Note that what matters here are the perceptible qualities of the object *as represented*. The object may be a merely *intentional* object, and the perceptible qualities of which a subject is aware may not be the real qualities of the object. A referee of this paper objected that, for someone with bad eyesight, viewing the world without one's glasses is a case of distinct phenomenology without any change in the perceptible qualities of the object. Nonetheless, there is most certainly a change in the perceptible qualities of the object *as represented* (as can be seen from the fact that the new experience is *informationally* impoverished compared to the experience with one's glasses on). The 'blurry view' reveals, as it were, a strange world of fuzzy objects (which one does not *believe in* of course). It will be important below to remember that introspection works on the world *as experienced* (indeed, how could it work on anything else). Thus, even in a totally hallucinatory experience there is still a 'world' to be experienced. If I ask what you are hallucinating you do not report that you are hallucinating a 'mental state' but rather, as it might be, a 'pink elephant'.

3. Is it possible to save the I-scanner theory with the claim that the 'perceptible' properties of intentional states are simply the contents of those states. Then we would face the problem which perhaps first exercised Hume: how do I know whether I am *believing* a proposition as opposed to merely *entertaining* it. The content is the same in both cases. Hume is characteristically straightforward: 'the difference between fiction and belief lies in some sentiment or feeling, which is annexed to the latter, not to the former, and which depends not on the will, nor can be commanded at pleasure. It must be excited by nature ... ' (1748/1962, §5, Pt. 2, p. 48). Of course, this is entirely implausible and the theory of introspection delivered below provides a much more reasonable account of our introspective knowledge of the intentional mental states. One might also note that the idea that we 'perceive' the *contents* of intentional states severely stretches the perception metaphor, perhaps to the breaking point, simply because there is no phenomenology of content-perception.

4. This 'advantage' depends upon the assumption that the problem of other minds does *not* stem from a genuine and fundamental asymmetry in the relation we bear to our own mind and the minds of others. Since I tend to think there is such an asymmetry, I welcome the fact that the theory of introspection defended below reintroduces the problem of other minds in its classical form. In fact, an obvious objection to the evidence theories turns the advantage on its head: the

evidence theory's denial that we stand in a relation to our own states of mind quite distinct from the relation we bear to the mental states of others offers a direct refutation of the evidence theory.

5. At least mental states of minds *capable* of self-knowledge – there is a serious, though for us tangential, problem lurking here with respect to animal minds. If having a mind itself entails that the subject possesses generally accurate self-knowledge then most animals will end up mindless. I take it that this is not a welcome result and thus requires some kind of hierarchy of minds, only some of which will be capable of introspection. It is not altogether clear that those of a transcendentalist bent can accommodate a scale of more or less primitive animal minds (see the difficulties that arise in, for example, Davidson 1982).

6. In anthropic cosmology it is always possible to appeal to pure chance: the fact that the entire universe is well structured to create intelligent observers might be a mere cosmic accident. In a way, that is still a positive account underwriting the anthropic discoveries. In a similar vein, I suppose that a transcendentalist about introspection could maintain that it is just a 'brute fact' that creatures deserving the label 'person' appeared on the Earth. Then the fact that nothing could *be* a person in the absence of introspective capabilities would be the most extensive account of introspection possible. But this is extremely implausible. It could be hardly a matter of chance that our brains support our abilities to achieve self-knowledge, for this would be a continual miracle, though I suppose it might have been chance events that structured a brain with those capabilities. Nor is there the slightest reason to regard the existence of the high level cognitive activities which underlie introspection as a 'brute fact' of nature (the more so if there is, as I believe, a very good positive account of introspection available).

7. This point is intended to be uncontroversial and not metaphysically or epistemologically 'deep'. Suppose I report that I am in pain and someone asks me how I know this. The obvious, and correct, answer is 'I feel it' (and of course I know what pain is don't I).

8. It is interesting to note, as a referee of this paper did, that certain elements of Wilfred Sellars's views are in some ways anticipatory of Dretske's theory. Recall that, in Sellars's famous 'myth of Jones' (Sellars 1956), Jones teaches his fellows to apply his new-fangled 'theory of mind' not only to others but also to themselves. However, Sellars never explains how this is supposed to work, and he explicitly suggests that the self-application of mentalistic concepts is 'non-inferential'. This emphasis is rather at odds with Dretske's 'displaced consciousness' theory. This sort of theory of introspection stems naturally from any representational theory of consciousness and so we find something similar sketched by Michael Tye (1995); see also Seager (in press, especially ch. 6). Some features of this theory also appear in David Rosenthal's 'higher-order thought' theory of consciousness (see Rosenthal 1986) and Daniel Dennett's 'quasi-eliminativist' account of consciousness (see Dennett 1991). Some commentators on Davidson's approach have also come close to the displaced consciousness view of introspection; see for example Holly 1986. A version of the view is also developed by Peter Carruthers 1996 in the context of the debate between the simulationists and theory-theorists about the nature of mental state attributions (this debate is briefly described below).

9. 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 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).

10. 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.

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

12. 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, along with many other mechanisms, that the theory of mind which grounds introspection is passed on to our children).

13. 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 both the I-scanner and evidence theories, Moore's paradox is somewhat troubling for, in the case of the former theory, surely I can 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 the I-scanner theory. In terms of the evidence theory, surely I could observe that my behaviour indicated that I did not believe some proposition that I could see was true. But contrary to the evidence theory it is clear that seeing that p is true *trumps* any behaviour I might observe in myself so far as my own beliefs about my beliefs are concerned (although self deceptive error is possible here so that others can, in some cases, know what I believe better than I do myself, but note that in such cases I will deny the *truth* of some proposition, not just the belief in it).

14. Let me emphasize again that one does not need to know that one is taking *p* to be true before one can attain the introspective knowledge that one believes that *p*. Simply consciously *taking p* to be true is sufficient to ground the application of the concept of belief (just as consciously *seeing* the tiger is sufficient to ground the application of the concept of seeing).

15. Notice also that in this case I am not coming to know that I like eggplant curry via an introspective knowledge of my preferences. Introspection of preferences is no different than introspection of desire; it depends upon the perception of the valued qualities of the options facing one (the curry, whether real or imaginary, *tastes* good – hence, given a pretty basic understanding of mind, I know I *like* it). There is an ambiguity of expression that we must be wary of. Animals, no less than ourselves, are said to 'know what they like'. This expression should not be interpreted as crediting animals with introspective knowledge of their preferences or desires. It means no more than that they can choose according to their desires or preferences. Introspection of desire and preference of course requires the possession of desires and preferences but crucially it requires all the other conceptual machinery of introspection as well. It is also worth remembering that human introspection becomes very complex, intellectualized and intricate as we bring to bear our cognitive powers upon our own self understanding.

16. The issues are complex here, but the 'compensations' that successful autistic adults make are clearly incomplete. Do they thus lack the requisite theory of mind? Not necessarily. The sort of evidence we use in our applications of our mentalistic concepts may well depend upon a set of specialized quasi-sensory neural systems. For example, Baron-Cohen (1995) stresses the importance of what he calls the 'Shared Attention Mechanism' by which people coordinate and direct their joint behaviour towards a particular object. Baron-Cohen claims that SAM is missing or weak in autistic children. We may conjecture that this deficit cannot be remedied by any amount of knowledge about the theory of mind. Of course, we must also bear in mind that the theory of mind is, despite its familiarity, very complex and the compensations that autistic adults make may well reflect an incomplete grasp of this complexity.

17. For a debate of the merits of theory-theory versus simulation with regard to self-knowledge see the papers by Robert Gordon and Peter Carruthers in Carruthers and Smith 1996.

References

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

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*, Cambridge, MA: MIT Press.

Baron-Cohen, S., A. Leslie and U. Firth (1985). 'Does the Autistic Child have a "Theory of Mind"?', *Cognition*, 21, pp. 37-46.

Bilgrami, Akeel (1992). *Belief and Meaning*, Oxford: Blackwell.

Carruthers, Peter (1996). 'Simulation and Self-knowledge: a Defence of Theory-Theory', in Carruthers and Smith 1996.

Carruthers, Peter and Peter Smith (1996). *Theories of Theories of Mind*, Cambridge, Cambridge University Press.

Churchland, Paul (1981). 'Eliminative Materialism and Propositional Attitudes,' *Journal of Philosophy*, 78, pp. 67-90.

Churchland, Paul (1995). *The Engine of Reason, the Seat of the Soul*, Cambridge, MA: MIT Press.

Davidson, Donald (1982). 'Rational Animals,' in *Dialectica*, 36, pp. 318-27. Reprinted in E. Lepore and B. McLaughlin (eds.) *Actions and Events: Perspectives on the Philosophy of Donald Davidson*, Oxford: Blackwell.

Davidson, Donald (1984). 'First Person Authority', *Dialectica*, 38, pp. 101-112.

Davidson, Donald (1987). 'Knowing One's Own Mind,' *Proceedings and Addresses of the American Philosophical Association*, 60, pp. 441-58.

Dennett, Daniel (1987). *The Intentional Stance*, Cambridge, MA: MIT Press.

Dennett, Daniel (1991). *Consciousness Explained*, Boston: Little, Brown and Co.

Descartes, René (1649/1985). *The Passions of the Soul*, in John Cottingham, Robert Stoothoff and Dugald Murdoch (trans./eds.) *The Philosophical Writings of Descartes (vol. 1)*, 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.

Holly, W. (1986). 'On Donald Davidson's First Person Authority', *Dialectica*, 40, pp. 153-156.

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

Lyons, William (1986). *The Disappearance of Introspection*, Cambridge, MA: MIT Press.

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

Rosenthal, David (1986). 'Two Concepts of Consciousness,' *Philosophical Studies*, 49, pp. 329-59.

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

Sacks, Oliver (1995). *An Anthropologist on Mars*, Toronto: Knopf.

Seager, William (in press). *Theories of Consciousness*, London: Routledge.

Searle, John (1992). *The Rediscovery of the Mind*. Cambridge, MA: MIT Press.

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.

Tye, Michael (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*, Cambridge, MA: MIT Press.