

*Philosophical Explorations*, Vol. 1 (1998), pp. 83-106.

### Self-Knowledge and the 'Inner Eye'

What is knowledge of one's own current, consciously entertained intentional states like? Is it like a sort of inner awareness? If so, what sort? In this paper I want to explore the prospects for a quasi-observational account of a certain class of cases where subjects appear to have self-knowledge. The ones that I have in mind are of the kind that preoccupied Descartes; the so-called *cogito*-like ones (Descartes, 1969). These are cases in which one is currently consciously thinking a thought with a given propositional content, say, the content, *water is transparent*, while thinking *about*, or reflecting *on* it, as when, for example, I think to myself, *I am currently, consciously thinking that water is transparent*. They are particularly interesting cases on which to focus attention for two reasons.

First, they are arguably the most plausible examples of *authoritative* self-knowledge; of knowledge whose possession by subjects gives them an epistemic advantage over others. Subjects occupy an epistemic position with regard to what they are currently consciously thinking which better places them over others to know what the contents of those thoughts are (Burge 1985, 1988, 1996; Heil 1988, 1992). This makes for an asymmetry between first- and third-person knowledge of subjects' intentional states. The asymmetry calls out for explanation in the light of the fact that subjects' knowledge of what they are currently, consciously thinking is typically not based on evidence (Davidson, 1984, 1987). Since beliefs that are not based on evidence are not generally thought to be more reliable than ones that are, it is puzzling why the non-evidenced based character of *cogito*-like thoughts should actually place subjects in a *better* position than others to know what thoughts they are currently, consciously thinking.

Second, and relatedly, the *cogito*-like cases are ones in which a subject is reflecting on, or thinking about, a thought *of which she is conscious*. The thought reflected upon is not only itself a conscious thought: it is a thought *about* which she is consciously aware. The *cogito*-like cases thus provide us with the best possible examples of thoughts concerning which one *might* have a sort of ‘inner’ awareness. If there were any prospects for a quasi-observational account of self-knowledge, the prospects would seem to be best for these cases.

My aim is to argue that the prospects for a quasi-observational account of the *cogito*-like cases of self-knowledge are much better than is typically assumed in current discussions of self-knowledge (Shoemaker 1994, McDowell 1994, Peacocke 1996, forthcoming, Wright forthcoming). I shall proceed in three stages. First, in section one, I shall provide a rationale for the claim that we need an epistemology of authoritative self-knowledge; specifically, of the *cogito*-like cases. In section two, I shall argue that contentful properties in the *cogito*-like cases of self-knowledge have two features that are also possessed by observational properties of objects in cases of perceptual awareness. Then, in section three, I shall develop a quasi-observational account of the *cogito*-like cases of self-knowledge by considering various accounts of the nature of observational properties, and specifically, of secondary qualities, and applying them to the *cogito*-like cases. I conclude by addressing some important objections to the account.

### I. Motivating the Account

My knowledge of what I am currently, consciously thinking is typically not based on evidence.<sup>1</sup> I do not normally go through a process of inference from my (first-order) thought, say, that water is transparent, to arrive at my (second-order) reflective thought that I am thinking that water is transparent. I do not use my first-order thought as a ground or reason for my second-order thought. My first-order thought and my second-order reflective thought about it are not normally mediated by some further thought or experience. I do not typically feel the need to justify my second-order

thought on the basis of my first-order one. Of course, it does sometimes happen that I am challenged by others, or am in a state of self-doubt about what I am currently consciously thinking. And in these cases, I might engage in a justificatory process. But these are not the typical cases.

Because my awareness of my current, consciously entertained thoughts is not evidence-based in these sorts of ways, it is puzzling why I should be assumed to be authoritative with regard to the contents of those thoughts. Why is it that I am in a better position than others to know what thoughts I am currently, consciously thinking? What is it about this position that confers upon me, but not upon others, this special epistemic right?

Is it simply that I think my thoughts whereas others do not? But how could this fact alone explain my favoured position with regard to knowledge of the *contents* of these thoughts? It may be true that one cannot think a thought without thinking its content. But *thinking* a content and knowing that a thought *has* that content are distinct matters. Moreover, since not all knowledge is authoritative, knowing authoritatively that a thought has the content it does is another matter still. It seems, then, that the authoritative position I occupy with regard to knowing the contents of my current, consciously entertained thoughts is not something that can be explained by the mere fact that I think them.

Perhaps there is something about the direct or immediate nature of my knowledge of my current, conscious thoughts that helps to explain the authoritative nature of that knowledge. The directness or immediacy of the relation between second-order thought and thought reviewed is a feature of *cogito*-like thoughts that has been noted by many (Burge 1985, 1988, 1996; Heil 1998, 1992). However, few have attempted to exploit it in an explanation of the authoritative nature of such knowledge. One reason for this is that an appeal to some kind of epistemic directness has been thought to commit one to a ‘Cartesian’ conception of the mind (McDowell, forthcoming; Wright, forthcoming).<sup>2</sup>

According to this, the mind is a kind of inner theatre, viewable by a kind of 'inner eye'. The immediate objects of one's thoughts are 'inner'; mental phenomena such as sensations, perceptual experiences and current, conscious thoughts. By attending to these so-called inner objects, one can know both one's own mind and what seems to be the case in the world beyond one's mind. Further, the existence and the nature, not only of one's sensations, but also of one's contentful thoughts, is independent of what may or may not exist beyond one's mind.

Typically, this conception of the mind is associated with a commitment to the view that a subject has privileged access to her own current, consciously entertained thoughts. This is so, not just in the sense that she is in a better position than others to 'view' them, and so to know them as the thoughts they are, but also in the stronger sense that her knowledge of such thoughts is either incorrigible or infallible.

Many find this conception and its associated commitments unacceptable (McDowell 1986, 1994, forthcoming; Wright forthcoming). There are various reasons for this, but one is particularly significant in the present context. This is a commitment to externalism in the philosophy of mind - to the view that subjects' intentional states have contents, some of whose natures are individuation-dependent on factors beyond the bodies of those subjects. Thus, for example, I cannot think a thought with the content, *tigers are feline animals*, in a world in which there are no tigers. This commitment forces a rejection of internalism in the philosophy of mind, the so-called Cartesian conception being a form of this.

Externalism is a metaphysical thesis concerning the nature and individuation conditions of contentful mental states (Burge 1979, 1986). It implies the falsity of the 'Cartesian' conception by rejecting the possibility that a thinker could think the very contentful thoughts she in fact thinks in her actual environment, in a world in which nothing whatever exists beyond the confines of her mind. The epistemology of self-knowledge associated with the 'Cartesian' conception seems incompatible with externalism because it seems to imply an internalist view of the nature and individuation

conditions of intentional content; a view according to which the natures of contentful intentional states are not individuation-dependent on factors beyond subjects' minds.

As this suggests, the 'Cartesian' conception of the mind has both metaphysical and epistemological aspects. The broadly observational account of self-knowledge it assumes performs a role in an epistemology of self-knowledge, in helping to make intelligible the asymmetry between self-knowledge and knowledge of others with regard to the immediate and authoritative nature of the former. However, in suggesting that contentful mental states compose an 'inner' realm whose contents are autonomous with regard to factors beyond the minds of subjects, the conception suggests more than just an epistemology of self-knowledge. It implies internalism with regard to the nature and individuation conditions of contentful mental states; metaphysical, and not just epistemological internalism.

Externalists who think that the 'Cartesian' conception of self-knowledge entails a corresponding internalist metaphysic will reject the broadly observational account of self-knowledge it assumes. However, the metaphysical and epistemological aspects of the 'Cartesian' conception may be, and I think are, separable. If so, then a broadly observational account of self-knowledge need not be worryingly 'Cartesian' in its metaphysical implications. Further, it may be that there is no better account to be given of the authoritative nature of subjects' thoughts about what they are currently consciously thinking than what a broadly observational account can offer.

I believe that a broadly observational account, based on the direct, immediate nature of a subject's knowledge of her own current, conscious thoughts, can ground the asymmetry between self-knowledge and knowledge of others. However, there are those who think that no such account is needed. So before proceeding any further, it is important to briefly consider these.

The phenomenon whose explanation appears to require an epistemology of self-knowledge is the asymmetry between self-knowledge and knowledge of others (Davidson 1984, 1987). Someone who is unsympathetic to an internalist epistemology

might respond to this apparent need in one or the other of two ways. On the one hand, she might maintain that the explanation needs no epistemology of self-knowledge (Shoemaker 1988, 1994). Self-knowledge arises because it is in the nature of first-order states to cause, but not to provide reasons for, second-order, reflective states about them. Subjects need no reasons, nor any evidence, on which to base their reflective thoughts about their current, conscious states.<sup>3</sup> To be a first-order contentful state just is to play a causal role in the generation of other mental states, including reflective ones, and behavioural responses, given environmental stimuli, in a subject. The directness, or immediacy of the relation between reflective thought and thought reflected upon is not an epistemic matter, but a causal-cum-metaphysical one. Reflective states comprise part of the functional roles of first-order states; and functional roles determine what it is to be such states.

On the other hand, she might maintain that although an explanation of the asymmetry between self-knowledge and knowledge of others does require an epistemology of self-knowledge, it does not require an internalist epistemology. What explains the asymmetry is simply the existence of a reliable causal connection between first- and second-order contentful states, not something that must be available to the knowing subject. It may not be in the *nature* of first-order states that they play a causal role in the generation of second-order ones. However, the existence of a reliable causal connection between such states is sufficient to explain self-knowledge and the asymmetry between a subject's knowledge of her own contentful states and others' knowledge of those states.

Both of these responses suppose that the existence of causal relations between first- and second-order contentful states is sufficient to explain self-knowledge and the asymmetry between it and knowledge of others. The important difference between the two is that, whereas in the latter, the relation between first and second-order state is causal and contingent, in the former, it is causal and non-contingent, since it is part of the nature of a first-order state to be apt to cause second-order states and to mediate between

environmental stimuli and behavioural output. In neither case, however, is it required that subjects themselves be in any special state of ‘inner awareness’ or to have access to the causal conditions that ground the existence and reliability of the reflective thought. That is to say, both responses reject epistemic internalism.

However, this is just what makes both responses deeply unsatisfactory. If the phenomenon whose explanation seems to require appeal to an observational account of self-knowledge were *simply* the phenomenon of self-knowledge, and not that of *authoritative* self-knowledge, these responses might be plausible. But the phenomenon that calls out for explanation here is the asymmetry between self-knowledge and knowledge of others. And that asymmetry is anchored in the presumption that self-knowledge is, *unlike* knowledge of others, authoritative.

So consider the reliabilist response. According to it, what matters to self-knowledge’s being knowledge is that there be a reliable causal connection between first- and second-order reflective state. Since, however, this need not be known to be the case by the subject herself, it seems impossible to justify the claim that subjects are in a better position to know what thoughts they are currently consciously thinking than others. For there is no reason to suppose that there is *not* a reliable causal connection between others’ knowledge of subjects’ first-order states. If not, then there is no asymmetry to explain. Reliabilism does not have the resources to provide a reasoned explanation for the fact that subjects are sometimes authoritative with regard to knowledge of their own contentful states, and that this is not an accident.

Consider, then, the first (functionalist) response. Here it is no accident that subjects possess self-knowledge, since it is part of the nature of first-order intentional states to be apt to cause second-order, reflective ones. So the relation between first-order intentional states (types) and second-order reflective states (types) is non-contingent. Still, there is a problem explaining the asymmetry between first and third person knowledge of a subject’s first-order contentful states. It is true that it is not part of the nature of my contentful first-order states that they are apt to cause reflective second-

order ones in others. This makes for a kind of *directness* between first-order and second-order states in me that is missing when others reflect upon my first-order states. But the directness has no epistemic clout. It is simply causal-cum-metaphysical. That I think the first-order thoughts I think explains why I think the second-order thoughts I think. But it does not explain why I am in a better position than others to know which contentful thoughts I am thinking. This is knowledge about a contentful thought; and to employ the concepts that are constitutive of that thought is not thereby to know that that thought has those concepts as constituents. Thinking a concept is not itself to reflect upon it. So the non-accidental character of self-knowledge does not by itself provide me with an epistemic advantage over others. What matters for this is not just that my reflective thought be non-accidental, but that it give me an epistemic *entitlement* to knowledge of my first-order contentful thoughts that others do not have (Burge, 1996).

So there is a need for an explanation of the asymmetry between self-knowledge and knowledge of others that can account for the fact that, in the former but not the latter, such knowledge is authoritative. And this need is not met by accounts that attempt to mark the asymmetry in causal terms that are not epistemically internalist. Let me proceed, then, to develop one.

## 2. Direct Epistemic Access and First-Person Authority<sup>4</sup>

Let's begin with the thought, noted by so many, that a subject's awareness of her own current, consciously entertained thoughts, is peculiarly direct. It is direct in being non-evidence based, non-inferential, non-justificatory, and unmediated by some further intentional (or sensational) state.<sup>5</sup> Call this kind of directness *epistemic* directness, and the awareness associated with it *direct epistemic access*. How does the fact that I may have direct epistemic access to my own current, conscious thoughts help to explain my favoured position with regard to knowing that those thoughts have the contents they have?

My view is that there are certain features of properties to which subjects have direct epistemic access that are both essential to them being directly epistemically accessible and also to mark them off from other sorts of properties to which subjects do not typically have direct epistemic access (Macdonald 1995, forthcoming). One is that such properties are epistemically basic in that they are the fundamental and favoured means by which knowledge of the objects that have them is obtained.<sup>6</sup> Another is that such properties typically *are* typically as they appear to be to normal subjects in normal circumstances. These two features are applicable to properties that fall into the broad category of observational ones, specifically, primary qualities such as being square, and secondary qualities, such as being red, of objects of perceptual experience. But they are also applicable to contentful intentional properties constitutive of first-order states when these are currently consciously entertained and reflected upon while thinking them. Let me briefly explain.

Consider properties other than contentful intentional ones where the notion of direct epistemic access is generally thought to apply. I know that the table visually present before me is brown, and that it is rectangular, and this knowledge is plausibly understood as being direct (although not baseless). One explanation of how I can know directly that the table is an instance of this particular shape property, or an instance of this particular colour property, is that the instance is presented to me *as* an instance of that property through my sense of sight. I perceive the instance *as* an instance of that property, and so no evidence is needed to come to know that it is an instance of that property.

This is not true of other properties. Water, for example, is an instance of the chemical structure  $H_2O$ , but this instance is not manifested to me as an instance of that property through one of my senses. In short, certain properties seem to be ones to which we have direct epistemic access because they are observational: whether objects are instances of them can be determined just by unaided observation of those objects.<sup>7</sup> This is not to say that one can know which observational property is being manifested to

oneself on any one occasion just by being presented with an instance of it. One must be capable of recognising another instance of that property as of that property when presented with it on another occasion, and this requires one to have mastery of the concept of the relevant property. This means that the notion of direct epistemic access is intentional: for one to have direct epistemic access to a colour property such as the property, *brown*, it is not sufficient that one sees an instance of that property. One must see it *as* an instance of that property.

Certain features of observational properties characterise their epistemic directness in a way that marks them off from other properties. One is that they are epistemically basic or fundamental to knowledge of objects that instance them. The point is not that grasp of the observational properties of objects necessarily constitutes knowledge of their true nature. Rather, it is that such properties are those by which objects that instance them are typically known in the first instance. Knowing an object through instances of certain properties and not others favours certain ones epistemically.

Another, crucial feature of observational properties is that they *are* in general as they appear to be when instances of them are presented to normal perceivers in normal circumstances. Again, this is not a point about the natures of the objects that instance the properties but about the nature of the properties themselves. The nature of water may be such as to have the chemical constitution  $H_2O$ , but this is compatible with water's instancing certain observational properties that are such that *they* are as they appear to be to normal subjects in normal circumstances.

Both of these features apply to mental properties, with the qualification that each person is alone the subject of her own intentional and sensation states, so that each person is the only subject to whom instances of her sensation and intentional properties appear in an epistemically basic way. That they apply to sensation properties needs little argument, since it is generally acknowledged that (a) sensations are known by their subjects on the basis of their sensation properties, and (b) the nature of a sensation property is constituted by how instances of it feel to its subjects.

So consider a sensation property such as the property, *pain*. It seems to be one to which we have direct epistemic access. When subjects are in pain, their pains are manifested to them *as* instances of the property, *pain*. That property is epistemically basic to the event that is my having pain. So I know my pains by knowing them *as* instances of that property. And my pains instance a property, namely, the property, *pain*, which *is* as it appears to me to be in normal circumstances.

On the face of it, intentional properties are not like this. One difference, to which I shall return, is that they seem to contain no phenomenal element whatsoever, unlike cases of sensation and perception of primary and secondary qualities. Further, even in the case of sensation properties, where one would expect the analogy to work better than in the case of intentional ones, the analogy with observational properties is imperfect, for two reasons. The first is that one's access to sensation properties does not appear to be through any medium of sense. There does not seem to be any fixed use of an *organ*, as there is with visual and auditory experiences. Of course, one could argue that many experiences, such as proprioception and kinaesthesia, involve no fixed use of an organ either; so that this isn't an absolute requirement on a property's being observational. Or, perhaps less plausibly, one could argue that there is a kind of 'inner eye' in the case of sensations too, which acts as a kind of internal scanning device (Hill, 1991). I shall return to this point in the final section of the present paper.

The second reason why the analogy is imperfect, however, is much more important. It is that observational properties are such that their possession by an object is importantly connected with their effects on normal perceivers. This is true both for the primary qualities such as that of being rectangular, where the connection between an object's being an instance of the property and how things look to normal observers in optimal conditions is thought to be *a posteriori* and contingent, and for the secondary ones, like that of being brown, where the connection between these and the best opinion of normal observers under optimal conditions is thought to be *a priori*, and further, thought by some to determine the nature of the property itself (Wright 19987, 1989,

1992). However, the sensation properties of which my pains are instances, although directly accessible to me, are not in general directly accessible in the observational way. For I, and only I, am the subject of my sensation states. Others may know what type of sensation state I am in by means of its effects in actions of mine, and so, for that matter, might I. But I *can* know them in ways that others generally do not.

Despite these qualifications, sensations properties are epistemically basic and epistemically direct in a way that makes them amenable to an observational model. Further, I think that these two features apply to intentional properties in the *cogito*-like cases in a way that also makes them amenable to an observational model. Consider the first feature. When one thinks of a first-order intentional state while undergoing it, from the point of view of a second-order intentional state, one's grasp of that first-order state is first and foremost a grasp of it *as* a state of a certain contentful type. The point is not that that state cannot be known by means of other properties (intentional or non-intentional); so it is not that I cannot think of that state apart from thinking of it as a state of a contentful type. It is that, when I *do* think about a first-order intentional state of mine, I typically think of it as a state of a contentful type.

Consider now the second feature: that such properties are in general as they appear to be to their subjects. This feature also applies to intentional properties in the *cogito*-like cases. The reason, which I shall develop more fully in section 3 below, is that the relation between them and normal subjects is in important respects like that between certain observational properties, namely, secondary qualities, and normal perceivers in normal circumstances.

Given that contentful intentional properties are in general as they appear to be to normal subjects in normal circumstances in the *cogito*-like cases, and given that subjects are the only ones to whom their contentful types appear when they appear in the epistemically direct way, it follows that, in these cases, one's first-order state cannot but be of the particular contentful type by which one grasps it. It could only be an intentional state other than one of the type a subject takes it to be in virtue of that

subject's grasping a different contentful property. But in that case one would be thinking a different thought altogether.

So the argument for authoritative self-knowledge based on a quasi-perceptual model consists of three premises:

1. A subject S typically thinks about her own intentional states *as* states of particular contentful types.
2. S's intentional states are of the contentful types that they appear to be.
3. No one *other* than S can be the subject of S's intentional states, so that when S thinks about her own intentional states as states of particular intentional types, S is the only one to whom those contentful types appear in this way.

Therefore, in general (that is, barring special cognitive failures)

4. S is authoritative with regard to the contents of her own intentional states.<sup>8</sup>

A number of points need to be made about this argument and the account of direct epistemic access upon which it is based. The first is that others can have epistemic access to my states. However, their access is evidence-based. It is not that they do not grasp my states, but that their grasp is in relation to my actions. Their mode of access is, as a result, irremediably evidence-based. This asymmetry is what explains the asymmetry between self-knowledge and knowledge of others that grounds first-person authority.

Second, when I know my states in this evidence-based way, there is no first-person authority for me. Third, when I know my states in an epistemically direct way, my knowledge is not only incorrigible (in that it cannot be shown by others to be false) but infallible (it cannot be false). This is so simply because I grasp the thoughts I grasp. The contentful type by which I grasp a first-order intentional state guarantees (but does not *make it the case*) that it is the state that it is. One cannot have a thought of a certain contentful type and misidentify it. Since to grasp a thought is to grasp its content, to grasp it as a thought of a different type would be to think a different thought altogether. In cases such as these, there is evidently no possibility of a contrast between what a subject is inclined to think, on the one hand, and what is actually the case, on the other - no possibility that one might think that one thinks a thought of a given contentful type and yet it fail to be the case that one thinks a thought of that contentful type.

### 3. Developing the Account

The argument for authoritative self-knowledge just given trades crucially on there being certain clear analogies between contentful properties in the *cogito*-like cases, and observational properties. Without this, the account of direct epistemic access that forms the basis of the distinctive authority subjects have with regard to certain of their own intentional states founders. The critical feature that observational properties possess, and whose possession by first-order intentional ones when consciously entertained helps to explain a subject's authoritative position with regard to knowledge of her own contentful states, is that they are in general as they appear to be to normal subjects in normal circumstances. This feature and the role it plays in the account of direct epistemic access does not by itself explain the authoritative status of self-knowledge, since it does not establish the asymmetry between self-knowledge and knowledge of others. But it forms part of that explanation, since what completes it is the further fact that (unlike cases of normal perception of observational properties) subjects are the only ones to whom their intentional states appear in the epistemically direct way.

The feature of observational properties to be accounted for and exploited in the account of direct epistemic access is that they are in general as they appear to be to normal subjects in normal circumstances. What kind of model of observational properties will best capture this feature in a way that is (a) applicable to cases of self-knowledge, and (b) compatible with metaphysical externalism, and so not worryingly ‘Cartesian’ in its metaphysical implications?

The claim that observational properties are as they appear to be in normal conditions to normal subjects is true of both the primary qualities, such as shapes, and the secondary qualities, such as colours. However, it seems to be true for different reasons. With primary qualities, the relation between the property and how it appears to be to normal subjects in normal conditions seems to be contingent. Secondary qualities, however, are often assumed to bear a *necessary* relation to subjects’ perceptual experiences and/or beliefs (Wright, 1988, 1993). This is thought to be important for the purposes of developing an account of authoritative self-knowledge based on analogies with observational properties for at least two reasons. One is that, in cases of authoritative self-knowledge, the relation between second-order thought and thought reviewed seems to be non-contingent (Shoemaker, 1994). Another is that such knowledge seems not only to be non-evidence-based, but also *a priori* (McKinsey, 1991, 1994; Boghossian, 1997); and at least some accounts of the nature of secondary qualities construe propositions expressing the relation between such properties and effects on normal perceivers in normal conditions to be knowable *a priori* (Johnston, 1992, 1993; Wright, 1988, 1993).

Two of the three accounts of secondary qualities that I shall briefly consider in this section assume at least this much. The third does not require it. According to this account, secondary qualities, while being *non-accidentally* connected with effects on normal perceivers in optimal conditions, are not necessarily so connected, in the sense of logical or conceptual necessity. For reasons that will become apparent, this third account is better able than the other two to provide the basis upon which a quasi-perceptual

account of authoritative self knowledge in the *cogito*-like cases might be developed compatibly with metaphysical externalism.

### 3.1 Extension-Reflection, Extension-Determination, and the Dispositional View

Crispin Wright (1988, 1993) distinguishes between primary and secondary qualities in terms of differences in our understanding of certain bi-conditionals, which he calls ‘provisoed equations’, associated with them. According to him, the general schema expressing the relation between primary qualities and their effects on normal perceivers is the same as that expressing the relation between secondary qualities and their effects on normal perceivers. It is this:

Under conditions  $C$ , for all subjects  $S$ ,  $Fx$  iff  $S$  believes that  $Fx$ . (Wright, 1988).

Alternatively, in Johnston’s (1993) terminology, it is a ‘basic equation’ whose general form is this:

$P$  iff for any  $S$ : if conditions  $C$  obtain, then  $S$  believes that  $P$  (Wright, 1993),

where  $C$  is a (substantial) specification of conditions such that, given them,  $S$ ’s belief that  $P$  ensures the truth of  $S$ ’s belief (and so  $P$ ). The substantiality requirement on the specification of conditions  $C$  is imposed in order to rule out trivial (‘whatever it takes’) elaborations of the basic equation.

In the case of secondary qualities, the provisoed equation and basic equation, respectively, take the followings forms:

For any subject  $S$ : if  $S$  were perceptually normal and  $x$  were presented to  $S$  under perceptually normal conditions, then ( $S$  would believe that  $x$  was  $F$  iff  $x$  was  $F$ ).

and, for a secondary quality such as *red*,

*x* is red iff for any *S*: if *S* knows which object *x* is, and knowingly observes it in plain view in suitable perceptual conditions; and is fully attentive to this observation; and is perceptually suitable and is prey to no other cognitive disfunction; and is free of doubt about the satisfaction of any of these conditions, then if *S* forms a belief about *x*'s colour, that belief will be that *x* is red. (Wright, 1993)

Similar equations hold for primary qualities.<sup>9</sup> However, Wright maintains that there is a critical difference between primary and secondary qualities, which is revealed by what he calls the order-of-determination test. The former are such that experiential effects on normal perceivers in normal conditions *reflect the extensions* of the relevant properties. The latter, however, are such that experiential effects on normal perceivers in normal conditions *determine the extensions* of the relevant properties. For Wright, this distinction between extension-reflection and extension-determination is the basis for the primary/secondary quality distinction. Primary qualities are such that their effects on normal perceivers in normal circumstances reflect but do not determine their extensions. Their extensions are otherwise determined, and so although there is a *reliable correlation* between their presence and their effects on normal perceivers, this correlation is contingent, not necessary. Such properties just happen to have effects on perceivers, and although this may be non-accidental in the sense that there might be natural laws that establish a nomological connection between these properties and their effects on normal perceivers, the natures of such properties are not determined by those effects.

Secondary qualities, in contrast, are such that their natures *are* tied to and determined by effects on normal perceivers in normal conditions. More precisely, they are such that their effects on normal perceivers in normal conditions do not merely

reflect, but determine, their extensions. The dependency relation between them and perceptual effects is asymmetric, a dependency of such qualities on their perceptual effects, but not vice versa. This relation is a necessary one. Wright also maintains that it is knowable *a priori*.

The dispositional view of secondary qualities differs crucially from the extension-determination view in just this respect, at least on some interpretations (Johnston, 1992, 1993). According to this, the nature of a secondary quality can be specified by the following schema:

The property  $F$  = the  $T$  disposition to produce  $R$  in  $S$  under  $C$ <sup>10</sup>

where a  $T$  disposition is some type of disposition (for example, a probabilistic one, an invariable one, and so on),  $R$  is the manifestation of the disposition,  $S$  is the site of the manifestation, and  $C$  are the conditions in which such a manifestation can occur. A secondary quality is *response-dispositional* when a schema of this form is true and three further conditions hold. First,  $R$  essentially or intrinsically involves some mental state or process. Second,  $S$  is some subject or group of subjects. And third,  $C$  are conditions under which subjects can produce the specified responses. For a secondary quality, such as red, the schema takes the following form:

The property red = the standardly realised disposition to look red  
to standard perceivers under standard conditions.

The dispositional view is committed to the claim that the relation between secondary qualities and their effects on normal perceivers in normal conditions is a necessary one, and also that it is knowable *a priori*. However, it does not appear to be committed to the view that there is an asymmetric dependency relation between secondary qualities and effects on normal perceivers in normal conditions. The reason

for this is that the account need not be committed to an analytically reductive analysis of secondary qualities. In rejecting this commitment, it need not be unduly concerned with circularity charges. Circularity in the account would be a problem if it were committed to an analytically reductive analysis, and in that case asymmetric dependency would be on the cards. But since the dispositional view need not be analytically reductive, it neither needs to be concerned with circularity charges nor needs to be committed to anything as strong as the claim that the natures of secondary qualities are determined by effects on normal perceivers in normal conditions (Johnston, 1993).<sup>11</sup>

Nevertheless, in being committed to the necessity of the relation between secondary qualities and effects on normal perceivers, the dispositional view is committed to a mutual or symmetric dependency relation of each on the other. For, according to the account, it is in the nature of a dispositional property to be effect-sensitive: this is what makes it a dispositional property. This is how the dispositional view marks the distinction between primary and secondary qualities.

What are the consequences of applying these models to the *cogito*-like cases of authoritative self-knowledge, in an attempt to explain why contentful properties of first-order states are in general as they appear to be when they appear in an epistemically direct way to their subjects? Consider first the extension-determination view. It has the consequence that the natures of contentful intentional states, like the natures of secondary qualities, are determined by their effects on normal perceivers in normal circumstances. More specifically, in the *cogito*-like cases, they are determined by what Wright calls ‘the best opinion’ of normal perceivers in normal conditions. For one to have a first-order thought of a given contentful type just is for one’s best opinion to be that one is having a first-order thought of that contentful type, given that the conditions are normal. So, in reflection, one determines one’s first-order state to be of the contentful type that it is. One’s authority with regard to knowledge of what one is currently, consciously thinking is explained by the fact that, in cases of self-knowledge, the subject is the *only* one in the position to have a ‘best opinion’: ‘best opinion’ here

just is the subject's opinion. One cannot misidentify one's own current, conscious, contentful thoughts. The reason is that in reflection one determines them to be the contentful states that they are.

However, from the fact that one cannot misidentify one's own contentful states in reflection it does not follow that one determines those states to be of the contentful types that they are. Reflection is, in one respect at least, an apt characterisation of the special relation subjects' second-order thoughts bear to their first-order ones. In physical reflection, say, in a mirror, under normal conditions, the object cannot be misrepresented. So the object reflected is as it appears to be. But the reflection does not determine the object to be what it is; it does not determine the object's nature.

On the extension-determination view, in the case of mental reflection, reflection does determine the nature of the thought reflected upon. This is perhaps not openly in conflict with metaphysical externalism, but it is plainly in tension with that view. For, according to it, the contents of first-order intentional states are determined to be what they are, *not* by subjects' best opinions regarding them, but by mind-independent factors external to subjects' bodies. This tension could be resolved by claiming that, in cases of reflective self-knowledge, the contents of subjects' *second-order*, or reflective thoughts, are determined by factors external to subjects' bodies, which in turn determine the contents of subjects' first-order thoughts. But this would require one to maintain that, in such cases of self-knowledge, one thinks a first-order thought (about, say, what is visually present before one) by thinking a reflective thought *about that thought*, thereby bringing it into being. This is not only implausible, but it conflates two bases for psychological ascriptions (Moran, 1994).

According to Moran, one can distinguish between theoretical-descriptive and prescriptive bases for psychological ascriptions. He invites us to consider a person who, in the process of wondering about her current conscious intentional states, asks herself the question 'What do I think about X?' This question can be interpreted in two ways: first, as a theoretical-cum-descriptive one about an attitude antecedently held; and second,

as a prescriptive one about what the subject *ought* to believe about X. Normally, these two ways of interpreting such a question interact with one another. However, Moran argues that they are distinct and that it may be possible to detach one from the other.

Moran's point is that reflective thinking is not, or not always, exclusively of the prescriptive form, where the question of rational interpretation, of having reasons by which to criticise, evaluate, and amend one's thoughts or beliefs, enters into the very process of first-order belief formation and 'guides' one's own first-order reasoning. There are descriptive elements in such reasoning, even where there are also prescriptive ones. Daydreaming may be a case of this kind.

If this is right, the suggested strategy for resolving the tension between metaphysical externalism and the extension-determination view of authoritative self-knowledge conflates these two bases for psychological ascriptions, allowing only for the prescriptive basis. In doing so, it undermines the intuitive basis for metaphysical externalism. That basis just is the intuition that whatever determines the first order thoughts to be what they are also determines the second-order thoughts to be what *they* are; and that this is best explained in terms of an asymmetric dependency relation of second-order thought contents on contents of first-order thoughts, whose natures are dependent on factors beyond the bodies of subjects. That the contents of one's perceptual thoughts should be determined, not by factors in the environment of a subject, but by a subject's reflective thoughts about those thoughts seems incredible.

Turn now to the dispositional view. This need not be committed to the asymmetric dependency of properties on effects on normal perceivers.<sup>12</sup> So its application to the *cogito*-like cases does not result in outright tension with metaphysical externalism. However, it is committed to a symmetric dependency between properties and their effects on normal perceivers. And this has the consequence that the natures of such properties are necessarily connected with and dependent on their effects on normal perceivers in normal conditions.

This fact creates a tension with metaphysical externalism. For that view takes the natures of first-order intentional states to be determined by mind-independent factors external to persons' bodies. That is to say, it presumes an *asymmetric* dependence of the contentful types of which first-order states are tokens on factors external to persons' bodies. And this is in tension with the view that the natures of such types are dependent on (even if not determined by) their effects in reflection on normal subjects in normal circumstances. The dispositional view implies that first-order states could not have the contentful natures they do if those contentful types were not disposed to produce certain effects (in this case, reflective thoughts) in normal subjects in normal circumstances. But externalism implies that the contentful natures of such states are quite independent of such dispositions.

For this reason, the dispositional view is also an inappropriate one upon which to base an account of direct epistemic access for the *cogito*-like cases that is not worryingly 'Cartesian' in implying metaphysical internalism. This leaves a third and final possibility to consider, the Simple View.

### 3.2 The Simple View

What is needed is an account of secondary qualities that explains why in general they are as they appear to be and which also respects externalist commitments regarding the individuation conditions of contentful states. More precisely, what is needed is an account that construes the relation between such properties and effects on normal perceivers as an asymmetric relation of dependence of the nature of the latter on the nature of the former but which is capable of explaining why such properties are at least non-accidentally as they appear to be. One possibility is known as the Simple View (Campbell, 1993). According to this, secondary qualities have three essential features. First, they are mind-independent properties of objects. Second, they are *grounds* of

dispositions of objects to cause experiences of them. (That is, they *have*, rather than *are* dispositions.) And third, they are properties whose natures are *transparent* to subjects in experience. The idea of the account is that secondary qualities are neither to be identified with dispositions (unlike the response-dependent, or dispositional account) nor to be identified with what the dispositional theorist may be happy to call the ‘grounds’ of such dispositions, namely, the categorical physical properties of objects (which, on some dispositional accounts, are taken to be the primary-quality realisers of such dispositions). The objection that such a view faces is that on it, secondary qualities either collapse into dispositional ones, or collapse into primary qualities (the so-called physical realisers of dispositions), with the consequence that there is no distinctive explanatory work for such properties to do.<sup>13</sup>

However, I think that this account, or something very much like it, is capable of being developed and defended in such a way as to avoid this objection. Further, so developed, the account can be seen to serve as a plausible model for the relation between first-order and second-order contentful properties in cases of self-knowledge.

What needs accounting for in the Simple View is the distinctive explanatory role that secondary qualities can play, which neither the dispositional view nor the physical releaser view can account for. Specifically, what needs explaining is how the Simple View can be true: how it can *both* be that secondary qualities, such as colours, are *response-independent* properties of objects *and* that such qualities are ones whose nature is *transparent* to us in colour experience. Evidently, the claim that such properties are not dispositional ones, but are rather *grounds of dispositions*, must play a crucial role in accounting for these intuitions. However, in the absence of any explanation of what it is for something to be a ground of a disposition (other than the primary quality physical releaser view), it is unclear how these intuitions are to be accounted for.

One plausible way of explaining how these intuitions could all be true is by appeal to a biological account of secondary quality experiences similar to the biological

account of perceptual content defended by Mohan Matthen (Matthen, 1988). Matthen argues that such an account is capable of explaining how it is that perceptual experiences can be sensitive to the distal surface colour of objects (not their primary quality releasers) in a way that treats colour as an objective (that is, response-independent) feature of such objects. If Matthen is right, a biological account of perceptual content can give substance to the claim that colours are grounds of dispositions of coloured objects to cause perceptual experiences of colour in subjects, where such experiences are sensitive, not to the physical releasers of colour, but to the surface colours of objects themselves, such surface colours being response-independent.

According to Matthen, one of the chief virtues of a biological account is that it can accommodate cases of ‘normal misperception’: misperception that results from the *normal* functioning of perceptual devices. Such cases are normal because they are due to the imperfections of the device and are not explicable either in terms of the device’s being *maladapted* to its environment or to the device’s *malfunctioning*. An example of this is our ability to perceptually discriminate the colours of surfaces of objects. Colour vision makes use of an indicator in the organism of surface colour that is imperfect but the best available, so that its use is advantageous overall for the organism despite occasional error.

Consider, for example, gradual changes in surface colour that are misperceived as changes in illumination. Here the misperception occurs because changes in illumination gradients are typically gradual, and it is against this background that surface colours, and changes in colours, are perceived. Against the background of relatively stable illumination conditions, changes in colour can be identified as changes in surface colour. However, in circumstances in which a colour change is gradual it will be misperceived as change in illumination.

Matthen calls these cases of normal error because, although the organism’s colour indicator is imperfect, it is functional for the organism. Functionality permits error in ways not easily accommodated by a straightforward causal account, even one

that appeals to the notion of normal causes of misperception. Since, in the case of normal misperception, it is *normal* (in a normative sense) for the device to misperceive, where this does not amount to normally (that is, typically) being caused to misperceive, a causal account is not the appropriate one to employ. Normal misperception here cannot be explained in terms of the normal (typical) cause of misperception, since the device does normally (typically) correlate correctly gradual changes in illumination gradient with changes in illumination, not changes in colour.

Central to a biological account is the notion of a function. There are three central features to such an account: variation of features within a group, selection of some of these variants, and variable transmission of the selected features to descendants (Macdonald, 1995). This introduces an etiological element into the analysis of function (Millikan, 1984, Neander, 1991a,b, Macdonald, 1992). A biological account of perceptual content has it that organisms' colour indicators are functional because they have been selected for. They have been selected for because ancestors that have had such indicators have reproduced more effectively than those without. The differential rate of reproduction is due to the indicators' being reliably connected to the environment. Organisms with these indicators, however imperfect they may be, have been better adapted to their environments than those without them.

One of the interesting features of this account is the way in which it exploits the cause-sensitivity of colour perception to the presence of surface colour in the environment of organisms. A biological account takes the causal history of a feature to be central to its having a function. To make sense in the case of such a feature as perception of surface colour, it needs to treat colour as a response-independent feature of the environment. This is because it treats colour perception, and perceptual content, as an *adapted* function of organisms' colour indicators, where such adaptation is grounded in the *correct* or *reliable* detection by ancestors of surface colours in the environment.

This idea of an organism's colour-detecting device getting it right, of *matching* the facts in the environment, is explicable on a biological account in a very different way

from the way in which it is explained either on a dispositional view or on an extension-determination view. This is because on it, a non-accidental relation is forged between the presence of surface colour and perceptual experiences (via natural selection) *for reasons that do not tie the nature of colour properties to colour experience*. Getting it right matters to survival, and for this reason features of organisms are selected for. It matters to the reliable identification of ripe plants and fruit, and to the recognition of a mate, that one identifies the surface colours of objects correctly. The stickleback's colour indicator may incline it to fly toward red postal vans and post-boxes as well as to appropriate mates, but its colour indicator is functional nonetheless because recognising a mate by its colour matters to its reproductive success, and so to the survival of those of the species with such indicators.

A biological account is therefore capable of explaining how it is that these indicators become attuned to their environment in a way that preserves the response-independent nature of colour properties. It is for this reason that a biological account is capable of giving content to the claim that colours are *grounds* of dispositions to produce colour experiences in us - that they *have*, rather than *are* dispositions. The connection between surface colour and colour experiences is non-accidental even though it may fall short of being conceptually necessary.<sup>14</sup> The sense in which there is a non-accidental connection at the level of functionality concerns descendants. A mechanism is in place that explains the non-accidental connection between the two that depends upon, but is not reducible to, the statistical-cum-causal relations relied upon by the dispositional account. Yet the connection this mechanism forges falls short of the connection required by a dispositional view. For that account requires at least regular causal connections between surface colours and colour experiences to ground the nature of the colours. But a biological account of colour experiences, in treating them as adapted functions, requires only sufficiently reliable causal relations between surface colours and colour experiences to ground the functionality of colour experiences. Thus, a colour experience may have the adapted function of representing surface colour even

though few such experiences do in fact have functional effects. Consider, for example, the functionality of the flight response of the deer or the zebra to variations in surface light in its immediate environment. Such a response is functional for the species because it correlates with the presence of predators nearby. It is functional despite the absence in many, or even most cases, of predators in the immediate environment. For, failure to respond to such variations in surface light in this way can be costly to the species on those occasions where there is a predator nearby.

A biological account, then, makes intelligible why the functions of colour experiences should be sensitive to the presence in the environment of surface colours. Colour experiences become functional because they are sensitive to the presence of surface colours; they acquire adapted functions. These adapted functions arise because the match between surface colour and colour experiences enhanced the reproductive capacities of ancestors. That is, ancestors whose colour experiences detected surfaces of colour were better adapted to the environment (that is, they produced more offspring). The environment *shapes* the development of the species, which allows these experiences to be connected to the environment in an adaptive way. Colour experiences *ought* to be connected to the environment even when, as it happens, they are not. They have acquired the adapted function of registering surface colours because of the cause-effect relations that have held between surface colours and colour experiences in ancestors. But they are not reducible to such relations.

So a biological account is capable of illuminating how at least two of the three central claims of the Simple View could both be true, namely, that secondary qualities have natures that are response-independent, and relatedly, that secondary qualities are grounds of dispositions to produce experiences of them (since colour experiences' acquiring functions depends on this). Does it allow for the *transparency* of colour experience? I believe so. What matters to a biological account is not that registering the primary quality realisers of colours is *not* functional for an organism (for it may be). It is that if it is, it is because registering such realisers *suberves* the function of registering

surface colours of objects. And there is good reason to think that it does. As Matthen, discussing Land's theory of the mechanism of colour vision, remarks,

Suppose that somebody were to say that Land's theory shows not that we see colour, but that what we have so far mistakenly supposed to be colour vision is, in fact, a system for detecting different sorts of brightness gradients in the (proximal) image incident upon the retina. This hypothesis is perfectly compatible with everything that Land says about the mechanism of colour vision as it now exists. Only when we take the evolutionary development of colour vision into account do the two ways of describing Land's theory turn out not to be factually equivalent. The functional schema gives crucial importance to the (supposed) fact that the ability to distinguish different sorts of brightness gradient in an image does not by itself confer an evolutionary advantage upon us. Rather, this discriminatory ability persists only because, as it happens, it happens to correlate closely with the ability to discriminate surface colour. (Matthen, 1988, p. 23)

Similar remarks apply to the suggestion that it is the physical realisers of surface colour to which our colour indicators are sensitive. That there is a mechanism in place that enables us to register, say, reflectance of coloured surfaces does not show that it is not functional to identify surface colour, since functions *need* mechanisms.

#### 4. Conclusion

I have argued that there is a way of developing the Simple View of secondary qualities in such a way as to make intelligible how its central claims, and in particular, its claim that secondary qualities are *grounds* of dispositions to produce colour experiences in organisms, might be true. This is significant for the explanation of the authoritative status of self-knowledge consistently with metaphysical externalism, since the Simple View is compatible with the externalist commitment to the individuation-dependence of intentional content on factors external to subjects' bodies. Specifically, the Simple View construes secondary quality perceptual experiences as asymmetrically dependent on secondary qualities, just as externalism construes intentional content as asymmetrically dependent on factors external to subjects' bodies. Compatibility with externalism matters because, failing it, the broadly observational account of authoritative self-

knowledge would, in being epistemologically internalist, imply *metaphysical* internalism; and one of the central aims here has been to provide a broadly observational account of authoritative self-knowledge that is not worryingly ‘Cartesian’ in its metaphysical implications. The application of the Simple View to the *cogito*-like cases can help to explain the authoritative status of this type of self-knowledge given externalism inasmuch as it can help to explain how subjects’ direct epistemic access to their first-order contentful states is compatible with externalism. I think that it does this, since the account of secondary qualities upon which the notion of direct epistemic access is based, namely the Simple View, takes secondary qualities to be individuated by factors that are external to subjects’ experiences of them. Yet this does not prevent them from being epistemically basic and also as they appear to be to normal subjects in normal conditions.

I conclude that a quasi-observational account of authoritative self-knowledge for the *cogito*-like cases is defensible; and that such knowledge *is* a form of inner observation. But does this require the use of a fixed organ, a sort of ‘inner eye’, analogous to the eyes, ears, nose, and so on, for ordinary perceptual experience? And if not, does it matter to the plausibility of a quasi-observational account?

Some have claimed that there is no fixed use of any organ in introspective self-knowledge, and that this is one reason for thinking that such knowledge is not a form of inner observation (Shoemaker, 1994). However, I do not see why we must accept this view. Suppose that we did in fact discover that some region of the brain was strongly linked to awareness of sensations or occurrent thinkings, and perhaps had other similarities to outer sense organs (for instance, distortions due to manipulation of that region). Then might it not be tempting to think of it as an organ of inner sense? If so, then it seems that the possibility cannot be dismissed on *a priori* grounds. All that follows is that we *do* not see any comparable organ to eyes and ears, and that is neither here nor there.

Suppose, on the other hand, that we were unable to discover any such organ of inner sense. Would this show that inner sense was not a form of perception? Not necessarily; for often in perception itself there is not a fixed use of a single organ. Consider a case where I am listening to a quartet and I want to hear the viola separately from the other instruments. What I do is look at the viola player's left hand and pick up the speed of the vibrato, and suddenly her sound emerges from the overall blend.<sup>15</sup> In this case ears and eyes are cooperating in a way I cannot separate, and it seems wrong to view such perception as the operation of a simple transducer.

This is a case where more than one organ is implicated in perception. But perhaps this is not the best case, since here there is the use of *organs* of outer sense, even if these are not simple transducers. However, the point seems clearer for cases of kinaesthesia and proprioception, where there is an awareness by a subject of her limbs and where there appears to be no transducer at all.

So I do not think that much hangs on whether there is or is not an organ of inner sense, a sort of 'inner eye'. The fact that we do not now see any comparable organ to eyes and ears does not show that there is no such organ; and if it were to turn out that we were unable to discover one, the analogy with perception would not thereby fail to hold, since in perception too there are cases where there appears to be no transducer at all.

So commitment to the claim that authoritative self-knowledge is a form of inner observation need not commit one to the claim that such knowledge involves the use of a kind of 'inner eye'. Still, the doubt that such knowledge is like observation lingers; and I think that the reason is that, associated with the organs of outer sense is a distinctive phenomenology. The eyes present us with visual experiences, and these experiences are enjoyed in a distinctive way; a way that differs markedly from the way in which auditory experiences, presented by the ears, are enjoyed, or the way in which olfactory experiences, presented by the nose, are enjoyed. In each case, there is a way in which such experiences are had, a way that is distinctive and whose distinctiveness seems to involve the operation of a distinctive transducer.

The doubt that talk of ‘inner eyes’ masks is that there is any distinctive phenomenology associated with authoritative self-knowledge, particularly in cases of intentional states such as thoughts about one’s thoughts, and so on. It is easy to imagine that there is a distinctive phenomenology associated with one’s sensations; but it is not easy to imagine that there is one associated with one’s *thoughts* about one’s sensations. It is more difficult still to imagine that there is a distinctive phenomenology associated with one’s thoughts about one’s thoughts; a distinctive way in which these thoughts are enjoyed.

Cases of experiential memory, or memory experience, provide at least some reason to think that some states that consist in a subject’s awareness of her own first-order states have an experiential, or phenomenological character (Owens, 1996; Martin, 1992). Here one enjoys an experience that is of the past and experiences it as of the past (Martin and Deutscher, 1966). This is a distinctive way in which the experience is had; and it is very different from normal perception precisely in being an experience as of the past.

However, one need not rest the case on this alone. I think that, in the *cogito*-like cases, it is plausible to suppose that there is a way in which reflective thoughts are had, and this is clearly indicated by the fact that, in such reflection, one can *attend to* one’s own current, conscious states. Suppose that I am daydreaming. I catch myself doing this, give myself a shake, and chide myself for letting my thoughts drift when I ought to be reading the book in front of me. Suppose that I am infatuated with someone. I continually catch myself thinking about him when I ought to be thinking other thoughts. In both of these cases, and in countless others, attention brings to the forefront thoughts that are present in consciousness. But in attending to them, in becoming *objects* of reflection, these thoughts are enjoyed in a distinctive way. There is something that it is like to catch myself daydreaming. There is something that it is like to attend to one’s craving for an ice cream.

So I think that, in the *cogito*-like cases, there is a distinctive phenomenology, enough of one to allay doubts about whether it is in any way appropriate to claim that authoritative self-knowledge is a form of inner observation. The phenomenon of attention, which also occurs in ordinary perception, is one that also occurs in the *cogito*-like cases. And it is this feature - not the operation of an organ - that marks the presence of a phenomenology in introspective self-knowledge.<sup>16</sup>

Cynthia Macdonald

University of Manchester

## REFERENCES

- Davidson, Donald, 1963: 'Actions, Reasons and Causes'. *Journal of Philosophy* 60, pp. 685-700. Reprinted in *Essays on Actions and Events*. Oxford: Oxford University press, 1980, pp. 207-25.
- Davidson, D. 1984: "First Person Authority". *Dialectica* 38, pp. 101-111.
- Davidson, D. 1987: "Knowing One's Own Mind". *Proceedings and Addresses of the American Philosophical Association* 60, pp. 441-58.
- Dennett, D. 1987: *The Intentional Stance*. Massachusetts, MIT Press.
- Georgieff, Nicolas, and Jeannerod, Marc, 1998: 'Beyond consciousness of External Reality: A "Who" System for Consciousness of Action and Self-Consciousness'. *Consciousness and Cognition* 7, pp. 465-77.
- Macdonald, C. 1995: "Externalism and First-Person Authority". *Synthese* 104, pp. 99-122.
- Macdonald, C. 1998: "Externalism and Authoritative Self-Knowledge". In C. Wright, B. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.
- Moran, R. 1994: "Interpretation Theory and the First Person". *The Philosophical Quarterly* 44, pp. 154-173.
- Moran, R. 1997: "Self-Knowledge: discovery, Resolution, and Undoing". *European Journal of Philosophy* 5, pp. 141-61.
- Peacocke, C. 1992: *A Study of Concepts*. Massachusetts: MIT Press.
- Peacocke, C. 1996: "Entitlement, Self-Knowledge and Conceptual Redeployment". *Proceedings of the Aristotelian Society*, 96, pp. 117-158.
- Peacocke, C. 1998: "Conscious Attitudes, Attention, and Self-Knowledge". In C. Wright, B. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds*. Oxford: Oxford University Press.
- Shoemaker, S. 1988: "On Knowing One's Own Mind". *Philosophical Perspectives* 2: *Epistemology*. California: Ridgeview Publishing Company.
- Shoemaker, S. 1991: "Qualia and Consciousness". *Mind* 100, pp. 507-24.
- Shoemaker, S. 1994: "Self-Knowledge and Inner Sense". *Philosophy and Phenomenological Research* LIV, pp. 249-314.
- Wegner, Daniel, and Wheatley, Thalia, 1999: 'Apparent Mental Causation: Sources of the Experience of Will'. *American Psychologist* 54, pp. 480-92.

See his 'Content Preservation', *Philosophical Review* 102, pp 457-88, 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society* 96, pp. 91-116, and 'Reason and the First Person',

in C. Wright, B. Smith, and C Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Basil Blackwell, 1998) pp. 243-70.

---

<sup>1</sup> See Heil (1988, 1992), Burge (1985, 1988), Davidson (1984, 1987, 1988), Wright (1989), and Alston (1971). Some, like Wright, emphasize the non-evidence-based character of such knowledge, whereas others, like Heil, emphasize its non-empirically evidence-based character. Alston gives an illuminating account of the different senses that might attach to the notion of direct access. He argues that the notion of directness that is relevant to self-knowledge is epistemic, not causal, and is explicable in terms of being non-evidence-based, where this is distinct from being non-inferential. Heil (1992) endorses this view. For a classic argument to the conclusion that epistemic relations need not exclude causal ones, see Donald Davidson (1963). And for models of perception that explicitly construe the relations involved in perception as both causal and epistemic, see Peacocke (1992) and McDowell (1994).

<sup>2</sup> ‘Cartesian’ is in inverted commas because it is a name for what is commonly taken to be a Cartesian position. To what extent this conception is faithful to the work of Descartes is not a matter that will be addressed here. Although my conception of a ‘Cartesian’ position confirms in broad outline to McDowell’s and Wright’s, it may not tally with theirs in all of its details.

<sup>3</sup> Thus, Shoemaker says,

...believing that one believes that P can be just believing that P plus having a certain level of rationality, intelligence and so on, so that the first-order belief and the second-order belief have the same core realization, then it will be altogether wrong to think of the second-order belief in such cases as *caused* by the first-order belief it is about. (1994, p. 289),

and

...if the self-ascription of such mental states is grounded on inner sense, *a la* the object-perception model, then the concepts of those mental states cannot be causal or functional concepts. By the same token, if one thinks that these concepts *are* causal or functional concepts, one has a reason for denying that the self-ascription of mental states under these concepts is grounded on inner sense, *a la* the object-perception model, and has a reason for welcoming independent reasons for opposing that model. (1994, p. 268)

<sup>4</sup> This section draws on material from Macdonald (forthcoming).

<sup>5</sup> Note that I do not here mean that such knowledge is *baseless*. If it were, there would be no need for an epistemology of self-knowledge, no need for an account of the kind that is here being offered. For more on this, see Wright (forthcoming) and McDowell (forthcoming).

<sup>6</sup> That is to say, typically, in getting to know what an object is, these are the first of its properties that one gets to know that it has.

<sup>7</sup> I take certain properties to be observational if whether objects are instances of them can be determined just by unaided observation of those objects. Whatever ‘unaided’ means here, it does not mean ‘unconceptualized’: it may be that I cannot know *which* observational property is being manifested to me through its instance simply by being presented with an instance of it, since I must be capable of recognizing another instance of that property as of that property when presented with it. For recent discussion of observational properties and predicates, see Wright (1987). See also Peacocke (1983) for discussion of conditions necessary and sufficient for a concept to be observational. Peacocke further elaborates the sense in which the two features I take here to be central to a property’s being observational are features of observational concepts.

---

<sup>8</sup> This replicates the argument in Macdonald (forthcoming). Unlike Wright (1998, 1992), who maintains that one's authority with respect to one's own intentional states consists in the fact that one's best opinion concerning those states' contentful types fix the extensions of those types, I maintain that one's authority consists in the fact that one cannot in reflection misidentify the object of one's reflection, and this is so because the nature of the thought reflected upon determines the nature of the reflecting thought.

<sup>9</sup> Thus, for example, for a shape property, such as being pear-shaped, the basic equation is:

$x$  is pear-shaped iff for any  $S$ : if  $S$  knows which object  $x$  is, and knowingly observes it in plain view from a sufficient variety of positions in suitable perceptual conditions, and is fully attentive to these observations, and is perceptually suitable and is prey to no other cognitive disfunction, and is free of doubt about the satisfaction of any of these conditions, then if  $S$  forms a belief about  $x$ 's shape, that belief will be that  $x$  is pear-shaped. (Wright, 1993, p. 80)

<sup>10</sup> Johnston couches his discussion in terms of concepts rather than properties, where a concept  $F$  is ...the core of a conception or cluster of beliefs *de dicto* about  $F$ s. A belief *de dicto* about  $F$ s is one that can be properly reported by using the predicate 'is  $F$ ' in an oblique context. (Johnston, 1993, p 103)

But see note 5 above.

<sup>11</sup> Wright discusses this issue of circularity in Wright, 1992.

<sup>12</sup> If Johnston (1993) is right, it *is* not committed to this.

<sup>13</sup> Smith (1993), for example, argues that the only plausible way to understand how the three features just mentioned can be consistently combined is to treat secondary qualities as categorical (and so as primary) qualities of objects. But he sees no reason to think that they are, for two reasons. First, there is no analogue of measurement of primary qualities for secondary ones, and so no way to sever the connection between the property and its appearances. Secondly, there is a problem about constitution, since the real nature of such properties is 'transparent'. In the case of primary qualities, we can understand how the measurement of, say, the size of something can be connected with the measurement of the sizes of its parts. Nothing similar can hold for Campbell's secondary qualities, since their nature is, according to his account, transparent (that is, brute).

<sup>14</sup> But see Neander (1991a), who argues that the concept of a biological function is tied to that of a functional effect. If this is so, then it may be a priori knowable that the function of a colour experience is to represent colour.

<sup>15</sup> I am indebted to Adam Morton for this example.

<sup>16</sup> I would like to thank Graham Bird, Graham Macdonald, Adam Morton, and Andrew Woodfield for valuable comments and discussion.