Introspection

CHAPTER 43

INTROSPECTION

CYNTHIA MACDONALD

IT is a bleak winter day. I am gazing out of the window as I sit at my desk. A friend looks over and asks me what I am thinking. I reply: 'I'm thinking that a holiday in the sun would be nice'. This avowal seems to be a report on a current mental state of mine, of which I am aware. I am not only thinking that a holiday in the sun would be nice, but know what I am thinking, and I seem to know it in a way that my friend does not. How do I know what I am currently thinking?

'Introspection' is a term used by philosophers to refer to a special method or means by which one comes to know certain of one's own mental states; specifically, one's current conscious states. It derives from the Latin 'spicere', meaning 'look', and 'intra', meaning 'within'; introspection is a process of looking inward.[1] Introspectionist accounts of self-knowledge fall within the broader domain of theories of self-knowledge, understood as views about the nature of and basis for one's knowledge of one's own mental states, including one's beliefs, desires, conscious thinkings, and sensations.[2] Theories of self-knowledge are motivated by the apparent need to account for a number of striking features of at least some such knowledge, which ordinary empirical knowledge, including knowledge of the mental states of others, is typically thought to lack. Knowledge of certain of one's mental states is said to be epistemically *direct* or *immediate* in some sense (for example, in being non-inferential and/or non-evidence-based), and so *privileged* and/or *authoritative*, perhaps in being *incorrigible*, or *infallible*, or *transparent* to oneself (or all three of these). Introspectionist theories attempt to account for some,

[1] Although, as McLaughlin (forthcoming) points out, this is misleading: introspectionists do not suppose that there is such an organ as 'the mind's eye', nor do they suppose that introspection is literally visual.

[2] This contrasts with another use of the term 'self-knowledge', to apply to knowledge of the subject of mental states, or the self, and its nature (see e.g. Shoemaker 1968; Evans 1982; Bermudez 1998). This use will not form part of our discussion.

or all, of these features by reference to a special method by which this knowledge is obtained.

There is considerable disagreement, however, amongst those working in the area of self-knowledge—even amongst introspectionists, as we shall see—whether such knowledge possesses the features just mentioned, and if so, to what extent. Partly by way of clearing the ground for discussion of the introspectionist position, therefore, I begin, in Section 39.1 below, by characterizing two classical but radically opposed positions on self-knowledge, one broadly Cartesian (Descartes 1641/1984) and the other Rylean (Ryle 1949). Whereas the Cartesian position takes certain self-knowledge to be distinctive in the ways described above, the Rylean one denies that self-knowledge is distinctive in any way at all. I set out some reasons that might be adduced against the Rylean position, delaying discussion of the Cartesian one until Section 43.3. In Section 43.2 I discuss some examples of 'deflationary' positions on self-knowledge: ones that take some such knowledge to be distinctive in at least some of the ways mentioned above (specifically, in being authoritative) but reject the view that there is a special *method* or way by which one obtains such knowledge. I give some reasons why these positions might be thought to be unsatisfactory. Finally, in Section 43.3 I revisit the Cartesian position along with other introspectionist positions distinguishable from it, and conclude by recommending my own position over the others.

Since the case for an introspectionist position is intuitively stronger for knowledge of one's own sensations than for knowledge of one's own propositional attitudes, even when these are currently consciously undergone, my focus in what follows will be, by and large, on how introspectionist accounts of self-knowledge of current conscious propositional attitudes fare.

43.1 Two Classical Positions

Philosophical accounts of self-knowledge are concerned with a wide variety of mental states, which can be loosely grouped under two main sorts: propositional attitudes and sensations (McGinn 1982). Propositional attitudes are states such as those of believing, desiring, or thinking that *p*, for some propositional content *p*, which consist in a subject's bearing an attitudinal relation to a propositional content (as does, for example, my belief that water is transparent). Sensations are states such as those of being in pain, having a reddish visual experience, and seeming to see a round red apple. States of this sort have a characteristic phenomenological, or felt, quality, which is typically thought to be an essential part of their nature. As the examples cited indicate, included in this sort are both bodily sensations and perceptual experiences, where the latter have intentional objects (i.e. objects that may or may not have actual existence).

Suppose, now, that I am 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. And suppose that my so-called second-order thought—my thought that I am currently consciously thinking that water is transparent—constitutes knowledge. What is the status of this knowledge?

Descartes held that we know some of our intentional states, namely those that we are consciously undergoing while we are thinking about them, in an epistemically direct and authoritative way.[3] His paradigm for this kind of knowledge was the *cogito*, which includes not just thoughts like *I am now thinking*, but ones like *I am thinking that water is transparent*. Descartes thought that he had this special kind of knowledge because of the special epistemic relation he bore to his thoughts while he was thinking them. Many have construed this epistemic relation as being at the very least *immediate*, in the sense of being *non-evidence-based* (see Alston 1971; Davidson 1984, 1987, 1988; Burge 1985, 1988; Heil 1988, 1992; Wright 1989).[4] This immediacy is thought to extend beyond knowing *that* one is in a given intentional state—say, a state of thinking—to knowing what the content of that state is—say, a state of thinking that water is transparent.

According to this picture my knowledge of what I am currently, consciously thinking is not based on evidence. 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. My first-order thought and my second-order reflective thought about it are not normally mediated by some further thought or experience which serves as evidence justifying my knowledge of my first-order thought. I do not use my first-order thought as a ground or reason for my second-order thought. 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.

This generates a puzzle. My knowledge of my current, consciously entertained thoughts is not based on evidence, and beliefs that are not based on evidence are not normally thought to be more reliable than beliefs that are so based. Why, then, is my knowledge of what thoughts I am currently, consciously thinking *authoritative*? What is it about this knowledge that confers upon me, but not on others, whose beliefs about my thoughts *are* evidence-based, this special epistemic right?

[3] See his *Meditations*, especially Meditation 2: 'Lastly, it is also the same "I" who has sensory perceptions, or is aware of bodily things as it were through the senses. For example, I am now seeing light, hearing a noise, feeling heat. But I am asleep, so all this is false. Yet I certainly *seem* to see, to hear, and to be warmed. This cannot be false; what is called "having a sensory perception" is strictly just this, and in this restricted sense of the term it is simply thinking' (1641/1984: 19).

[4] Some, like Wright, emphasize the non-evidence-based character of such knowledge, whereas others, like Heil, emphasize its non-empirical-èvidence-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 the view that the notion of directness is epistemic, not causal. Gertler (2003) distinguishes two senses of 'direct' in 'direct access', one epistemic (what is here characterized as 'non-inferential') and the other metaphysical (here characterized as 'unmediated').

The mere fact that I think my thoughts whereas others do not cannot be the answer, true though this might be, since it does not by itself explain my favoured position with regard to knowledge of the *contents* of these thoughts. It may be 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. Further, since not all knowledge is authoritative, knowing *authoritatively* that a thought has the content it does is another matter still.

The Cartesian view is that it is the epistemically direct or immediate nature of the relation between the subject of the second-order thought and the first-order thought reviewed in thinking that second-order thought which explains the authoritative nature of *cogito*-type thoughts. This view is associated with a particular conception of the mind, according to which it is a kind of inner theatre, viewable by a kind of 'inner eye' (McDowell 1986, 1998; Wright 1998). The immediate objects of one's thoughts are 'inner': mental phenomena such as sensations, perceptual experiences, and current, conscious thinkings. 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.

The Cartesian conception of the mind is associated with a commitment to the view that subjects have privileged access to their own current, consciously entertained thoughts. This is so not just in the sense that they are 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 their knowledge of such thoughts is either incorrigible, in that they cannot be shown to be mistaken about them, or infallible, in that they simply cannot be mistaken about them, or both.

For the Cartesian, the special status of such thoughts that derives from the immediate, non-evidence-based relation between subject and thought makes for a kind of *transparency* of the thought reviewed to the reviewing subject, but not to others, where a transparent thought is one that is both infallible and certain, or indubitable, in that it is impossible for there to be any grounds for doubting that one has it and what its content is (Alston 1971). This transparency is due to the special method by which a subject has access to the contents of her own intentional states, which differs from others' access to those contents, and this is what confers an epistemic advantage on the subject in the sense of better placing her to know what the contents of those states are. The Cartesian account is thus a sort of introspectionist account of self-knowledge, what we might call a direct observational one (which we will revisit in Section 43.3).

This view contrasts sharply with the Rylean one. For the Cartesian, the special status of self-knowledge is a datum that serves as a secure basis on which to construct a theory of knowledge in general. For the Rylean, however, self-knowledge does not have any distinctive status. Unlike the Cartesian, the Rylean denies that there are features peculiar to self-knowledge that need accounting for by any theory

of self-knowledge (as distinct from a theory of knowledge in general). As Ryle himself puts it:

> The sorts of things that I can find out about myself are the same as the sorts of things that I can find out about other people, and the methods of finding them out are much the same. A residual difference in the supplies of residual data makes some differences in degree between what I can know about myself and what I can know about you, but these differences are not all in favour of self-knowledge. In certain quite important respects it is easier for me to find out what I want to know about you than it is for me to find out the same sorts of things about myself. In certain other important respects it is harder. But in principle, as distinct from practice, John Doe's ways of finding out about John Doe are the same as John Doe's ways of finding about Richard Roe.
>
> (1949: 155–6)

What 'privileges' self-knowledge over knowledge of others, when it does, is just the fact that we are more often able to observe our own behaviour than we are able to observe the behaviour of others, and so we have more data of the same kind (rather than data of a different kind) on which to base such knowledge.

Ryle's main objection to the view that self-knowledge is distinctive is to the idea of privileged access, or a special way of knowing one's own inner states, of the kind presumed by the Cartesian; namely, a sort of 'inner' perception of one's mental episodes. His objection is that if self-knowledge involved a second-order perception-like awareness, or consciousness, of a first-order state, this second-order awareness would itself have to be conscious, in which case it would itself require the subject to be aware of *it* by means of a 'third-order' awareness of the second-order state, and so on into infinity. However, Sosa (2003) claims that this argument rests on a confusion between a conscious episode (such as my consciously thinking that water is transparent) and being conscious *of* that episode (as I am when I am reflecting on my conscious thinking that water is transparent). I can be in a state of consciously thinking that water is transparent without being in a state of being conscious *of* that state. So I can be in a state of consciously thinking that I am thinking that water is transparent without being in a state of being conscious of that state.[5]

The Rylean alternative to the Cartesian model is to construe self-knowledge as a matter of bringing one's future behaviour into step with one's present behaviour in such a way as to make for a 'fit' between the two. One does this by being in a 'frame of mind' or 'being prepared' to say or do things in future circumstances by 'being alive' to, or observing, what one is now doing (Ryle 1949: 179). So self-knowledge requires

[5] Of course, according to higher-order theories of consciousness this is not a confusion, since it is definitional that 'conscious state' means 'mental state that one is aware of being in'. See, for example, Lycan (1996, 2003), who claims that 'a state of a subject, or an event occurring within the subject, is a conscious state or event, as opposed to an unconscious or subconscious state or event, iff the subject is aware of being in the state or hosting the event' (2003: 3). Higher-order theories will be touched on briefly in Section 43.3. As Sosa (2003) notes, Ryle does have another main objection to the Cartesian model, which is that self-knowledge does not always rest on introspection, but this, even if true, does not show that self-knowledge does not sometimes rest on introspection, and this is all that the Cartesian model needs in order to be vindicated.

observation of one's own behaviour.[6] Since one can 'be alive' to what others are now doing by observing their behaviour (even while being in a different frame of mind from them), there is no difference in kind between self-knowledge and knowledge of others. Ryle gives an example of a lecturer and a listener. The lecturer is alive to what she is doing, and prepared for the parts of the lecture she has not yet given, but her frame of mind is different from that of the listener, since she is creative while the listener is receptive. Still, the listener is alive to what the speaker is doing.

A neo-Rylean version of the position (McGeer 1996), which goes further in attempting to account for the authoritative status of self-knowledge, construes self-knowledge as a matter not just of there being a kind of fit between what one avows when one says 'I believe that *p*' and what one goes on to do or say, but of a subject's actually ensuring that there is that fit, by *making* her future actions and verbal behaviour fit her avowals. In that case, the truth makers for one's avowals are not one's present 'inner perceptions' of ones present inner states, but the future deeds that follow the words, which one brings about.[7]

The Rylean position is undermined by the fact that it seems so obviously true that, in at least some cases, there is an epistemic asymmetry between self-knowledge and knowledge of others and that this is due to the fact that subjects often know what they are thinking when they make avowals without appeal to evidence from their own behaviour as well before engaging in any such behaviour (Boghossian 1989), and by the fact that the arguments attacking the Cartesian alternative, which give support to this apparent truth, are not compelling.[8] The Rylean account may fit some cases of self-knowledge, specifically cases of knowledge of one's own non-episodic mental states, such as one's standing beliefs, particularly ones about which one is self-deceived, but it is the episodic ones, like the *cogito*-type cases, knowledge of which seems peculiarly authoritative. The neo-Rylean attempt to rescue authority while maintaining the central claim that avowals are not reports on inner states that underlie and cause behaviour by claiming that one's ability to ensure that one's future behaviour fits with and makes true one's avowals suffers from these problems as well as from problems of its own. It has the consequence, for example, that one cannot explain why people make the avowals they do *when* they do, as well as allowing almost any first-order belief which one can make true by undertaking a course of action, such

[6] As he puts it: 'Our knowledge of other people and ourselves depends upon our noticing how they and we behave' (1949: 181).

[7] Thus, McGeer says:

> The view I propose involves putting special emphasis on our own *agency* by recognizing that we are *actors* as well as observers and so can be good, even excellent, 'predictors' of our future behavior because *we* have the power to make these 'predictions' come true. Put simply, we are able to *ensure* a fit between the psychological profile we create of ourselves in first-person utterances and the acts our self-attributed intentional states are meant to predict and explain simply by adjusting our actions in appropriate ways.
>
> (1996: 507)

[8] As Davidson puts it, 'Ryle was wrong. It is seldom the case that I need or appeal to evidence or observation in order to find out what I believe; normally I know what I think before I speak or act. Even when I have evidence, I seldom make use of it' (1987: 441).

as my belief now that the balloon will pop (which I can make true by undertaking a course of action involving a dart), to count as authoritative (Brueckner 2001).

43.2 Some Deflationary Accounts

One might agree with the assessment of the Rylean position and with the view expressed by the Cartesian that at least some self-knowledge is distinctive, even to the extent of agreeing that the *cogito*-type cases are cases of this kind, but still reject the view, to which the Cartesian is committed, that what makes such knowledge distinctive has to do with a special method or way in which it is obtained. One might instead maintain that self-knowledge is no special cognitive achievement: it is knowledge, and it is authoritative, but it is not arrived at in a special way (or, indeed, in any 'way' at all).

There are a number of different positions that are of this general kind, and here we can only canvass a few. One stems from considerations about the nature of interpretation (Davidson 1984, 1987, 1989). According to this, there is an asymmetry between self-knowledge and knowledge of others because the grounds for self-ascriptions of beliefs differ from the grounds for other-ascriptions of beliefs. One argument for this stems from considerations concerning the nature of interpretation. When another interprets my speech, she must work out, from the external conditions in which my utterances occur, both what my utterances mean and what the contents of my beliefs expressed by those utterances are. But I do not need to work out what my utterances mean; when I say 'The balloon is red' I know what I mean by this. This knowledge is not based on evidence, observation of my own behaviour, or inference. Further, when another sets out to interpret my utterance, she needs to work out whether I hold the sentence uttered true; that is, whether I believe that it is true. My utterance itself along with other factors external to me provide defeasible empirical evidence for whether I do hold that sentence true. Unlike my interpreter, however, I do not need to rely on this evidence to determine whether I hold that sentence true. So, just as I do not have to work out what I mean by my utterance, I do not need to work out whether I hold the sentence uttered true. But if I know what my utterance of that sentence means, and I know that I hold that sentence true, I know what I believe.

One response to this argument is that it does not explain how I know that I hold a sentence true, that is how I know that I believe that it is true, and so does not explain how I know what I believe. Without an explanation of this, the argument seems to presume rather than explain authoritative self-knowledge (Gallois 1996). Davidson (1991) gives a different argument for the position, however, which doesn't depend on the nature of interpretation. According to this, although an interpreter must work out, from the external conditions in which my utterances occur, both the meanings of my utterances and the contents of my beliefs, I do not have to do this. But since whatever determines the contents of my beliefs also determines the contents of my beliefs about my beliefs, I cannot be mistaken about what I believe. My authority with

regard to my beliefs about my own beliefs and other mental states consists in the fact that such beliefs are infallible: whatever I think I believe I know I believe.[9]

In a similar vein, Burge (1988, 1996, 1998) argues that the authoritative status of certain self-knowledge—the kind of knowledge involved in the *cogito*-type cases—is due to the fact that they are contextually self-verifying. In his opinion, the proper way to view the *cogito*-type cases is as ones in which one's first-order thought is literally contained within, or is a constituent of, one's second-order thought, which is contextually self-verifying for the reason that in thinking the reflective thought one *makes* that thought true. One makes it true because in thinking the reflective thought one brings into being the first-order thought upon which one is reflecting. The relation between first- and second-order thought is non-causal, at least in part because there are not here two separate acts of thinking: there is no first-order thought, considered as a state distinct from the second-order, reflective one, to serve as one of the relata of the causal relation.

Although there are differences between these two versions of the deflationary position, both are committed to the view that certain self-knowledge is authoritative just because one and the same content is the content of both the first-order thought or belief and the content of the so-called second-order thought or belief about the first-order one. This being so, it is impossible that one should be in error about what that content is. One cannot have a thought of a certain contentful type and misidentify it. Since to think a thought is to think its content, to think of it *as* a thought with a different content would be to think a different thought altogether.

Some may object to this position on the grounds that it relies on a particular conception of what is involved in thinking second-order thoughts about one's first-order thoughts, but this conception is false because all second-order thoughts or beliefs about first-order ones involve *distinct* thoughts/beliefs. However, this would not by itself show that, in the *cogito*-type cases, such thoughts are not contextually self-verifying, since this feature of such cases might be explained differently. For example, it might be maintained that a necessary condition of thinking a reflective thought with a given content is that that content refers to a first-order thought content, and that a necessary condition for this is that the contents of the two thoughts are of the same type. If so, then although it would be true that thinking such a second-order thought suffices for its being true, and that thinking it makes it true, this would not be because there is no first-order thought content distinct from the reflective one to which it is causally related.

[9] Davidson typically couches this argument in terms that speak of presumptions of knowledge of meaning and the like, but, as the second of his arguments clearly indicates, there is more to the position than the presumption of authority. Here is a passage where the argument makes no use of it:

> An interpreter must discover, or correctly assume on the basis of indirect evidence, what the external factors are that determine the content of another's thought; but since these factors determine both the contents of one's thought and the contents of the thought one believes one has (these being the same thought), there is no room for error about the contents of one's own thought of the sort that can arise with respect to the thoughts of others.
>
> (1991: 196)

There is, however, another, deeper problem with this type of position. Even if we agree that certain cases of self-knowledge are special because their contents are infallibly known, and even if we agree that this is due to the fact that there is just one content thought and thought about, the authority awarded to their subjects by this means alone does not seem to have much to do with epistemology. In one clear sense of 'authority', someone whose knowledge is authoritative is in a *better epistemic position*, or is epistemically *better placed*, or better justified, than another to claim to have that knowledge. But this deflationary position does not award a subject's self-knowledge authoritative status for this reason. The authority here is simply to do with the fact that a subject of a second-order, complex thought, which includes as a constituent a first-order thought, thinks a thought with a single content. It is enough for authority that the thinker simply *thinks* that thought, since thinking it makes it true.

There is a further question whether, on this type of account, self-knowledge counts as a form of knowledge at all. It has been argued that in order for a mental state to count as a genuine form of knowledge it must be a genuine cognitive achievement—something that a subject can strive, and fail, to possess (Wittgenstein 1953; Wright 1989, 1998; Moran 1997, 2001). But this type of deflationary position does not allow for that.

Davidson and Burge develop their positions on self-knowledge against the background of a common presumption. Both are externalists with regard to the contents of propositional attitudes. Briefly, externalism is the view that certain intentional states of persons have contents that are 'world-involving' in that they depend on the existence of objects and/or other factors beyond the bodies of their subjects (Putnam 1975; Burge 1979, 1985, 1988; Davidson 1984, 1987, 1989—though Burge's view specifically concerns anti-individualism). Externalism has been thought to compromise the presumption of authoritative self-knowledge in virtue of its commitment to the view that the empirically discoverable determinants of a subject's thoughts might be better known to be what they are by another who has better knowledge of those determinants. Suppose, to use a well-worn example, that in the distant galaxies there is a planet, Twin Earth. Twin Earth is a duplicate of earth in every way, physically and phenomenologically described, apart from this: the substance that fills lakes, rivers, and streams and has all of the phenomenological qualities of water is constituted not by H_2O but by a different chemical structure, XYZ (Putnam 1975). Now consider a situation on earth in which a subject, who knows nothing about chemistry and Twin Earth, believes that she is currently thinking that water is transparent. Her justification for this seems weaker than that of another individual who, knowing about chemistry and Twin Earth, knows that the water content in question relates to H_2O, and that the subject has not (unbeknownst to her) been transported to Twin Earth. This seems to be a case where externalism conflicts with authoritative self-knowledge, intuitively understood.

The deflationary position taken up by Davidson and Burge responds to this threat by maintaining that since whatever determines the (water) content of the subject's first-order state to be what it is also determines the content of the second-order state,

and since the content of the first-order one just is the content of the second-order one, error with respect to the content of the first-order thought is not possible. While this may explain why the subject's thoughts about her water thoughts on earth are infallible, and also why her Twin's thoughts about her (water*) thoughts on Twin Earth are infallible, it does not really respond to the worry expressed by the above situation. That concern has to do with the fact that another, who has knowledge that the subject does not about both earth and Twin Earth, and water and water*, is better *justified* in her knowledge that the subject is indeed thinking that *water* is transparent. And she is better justified because she has, whereas the subject does not, *discriminative* knowledge—that the subject is thinking a water, rather than a water*, thought. Switching cases, where the subject has unwittingly been transported from earth to Twin Earth, are intended to illustrate the force of this threat. In one such case a subject is told that she has been transported to Twin Earth during sleep but is not told when the switch occurred, so that when asked which thought she was thinking a year ago she has no idea, even though she suffers no memory impairment (Boghossian 1989).[10]

Another version of the deflationary position that presumes the truth of externalism is held by McDowell (1986, 1998). Unlike the version just discussed, this one does not attempt to explain what makes for authoritative self-knowledge. On the contrary, it takes a sceptical stance on the view that there is something problematic about the claim that subjects have authoritative self-knowledge—indeed, infallible self-knowledge—that needs explaining. According to it, this is an illusion, whose source is the 'fully Cartesian' conception of the inner, which construes the mental domain as autonomous with respect to the world beyond the mind to the extent that a subject could think exactly the thoughts that she thinks in the world as it is

[10] It is not clear that this case shows that Burge's account is defective, unlike the example given in the paragraph immediately preceding this one. Burge (1998) has responded to it by maintaining that authoritative self-knowledge does not extend to cases of discriminative knowledge, and this is particularly so where memory is involved:

> I have maintained that the individual may not know whether yesterday he had an aluminum or twaluminum thought. He does not have discriminative knowledge of this form. But memory need not work by discrimination; it can work through preservation. The memory need not set out to identify or pick out an •aluminium rather than a twaluminum thought, trying to find one by working through the obstacles set by the switches. Preservative memory normally retains the content and attitude commitments of earlier thinkings, through causal connections to the past thinkings. That is one of its functions—maintaining and preserving a point of view over time. It need not take a past thought as an object of investigation, in need of discrimination from other thoughts. Memory need not use the form 'Yesterday I was thinking a—type of thought', where the memory attempts to *identify* the thought content as an object. Again, if it did, the individual might perhaps err by using a thought appropriate to the second environment in making an attribution to a thought event in the first environment. . . . The memory need not be *about* a past event or content at all. It can simply link the past thought to the present, by preserving it.
>
> (1998: 357)

It might be thought that the example in the paragraph immediately preceding this in the text makes essential use of the notion of discriminative knowledge, since the authoritative status of the subject's self-knowledge seems to be compromised by another's better knowledge, not only of the facts of chemistry, but also of Twin Earth and XYZ. But the point can be made equally effectively in terms of another's better knowledge of the facts of chemistry, without appeal to knowledge of Twin Earth.

• Q1

in a world in which there is nothing beyond her mind. Once the source of the so-called problem is exposed and the Cartesian grip is dislodged, one is free to view the inner and outer domains as 'interpenetrating' by construing the infallibly knowable appearances expressed by avowals of the form 'It seems to me that I am thinking that that F is G' disjunctively, as constituted either by the fact that I am indeed manifestly thinking that that F is G or by the fact that that merely seems to me to be the case.[11]

A question that arises is whether this view can account for authoritative self-knowledge any better than does the Davidson/Burge view. Suppose that I am in a situation in which I am thinking a singular *tiger* thought—say, I am thinking that that tiger has stripes—but am ignorant of the facts of biology and Twin Earth. On the present view I can have infallible knowledge that it seems to me that I am thinking that that tiger has stripes consistently with externalism because what I infallibly know is *either* that I am indeed manifestly thinking that that tiger has stripes *or* that that merely seems to me to be the case, and this is compatible with another person's better knowledge of which of the two disjuncts is in question. But in this situation another has this knowledge because she has discriminative knowledge that I do not—knowledge about the facts of biology and Twin Earth, where there are no tigers but only pligers (creatures that have all the phenomenological properties of tigers but differ in their biological constitution), knowledge that the *tiger* content in question relates to tigers rather than pligers, and knowledge that I have not (unbeknownst to me) been transported to Twin Earth. Because she knows this, she not only knows which disjunct is in question but knows the fact, disjunctively construed, that I know. So there does not appear to be any epistemic asymmetry between her and me. Further, and more importantly, she seems better *justified* than I am to know, when it seems to me that I am thinking that that tiger has stripes, that what it seems to me that I am thinking is either this: I am indeed manifestly thinking that that tiger has stripes, or this: that that merely seems to me to be the case.[12]

Finally, there is a 'neo-expressive' version of a deflationary position canvassed by Dorit Bar-On (2004), whose aim is to develop a view on avowals that respects two features that are naturally associated with them: (1) that they are semantically continuous and truth-conditionally interchangeable with other unproblematic statements, and (2) that they exhibit epistemic asymmetries with these other unproblematic, truth-conditionally equivalent statements. Her view is that avowals

[11] Thus, McDowell says:

> Short of the fully Cartesian picture, the infallibly knowable fact—its seeming to one that things are thus and so—can be taken disjunctively, as constituted either by the fact that things are manifestly thus and so or by the fact that that merely seems to be the case. On this account, the idea of things being thus and so figures straightforwardly in our understanding of the infallibly knowable appearance; there is no problem about how experience can be understood to have a representational directedness towards external reality.
>
> (1986: 150)

[12] Again, it might be claimed that authoritative self-knowledge does not extend to cases of discriminative knowledge and that this counts as one such case (see n. 10), but then the point can be put in terms of another's better knowledge of the facts of biology, without appeal to knowledge of Twin Earth and pligers.

are self-ascriptions of, but not judgements about, one's mental states, events, etc., and are expressions of one's mental states, but not reports on them. This combination of views allows her to maintain that avowals are both true and epistemically more secure than other self-ascriptions and other-ascriptions.

Bar-On recognizes that there is an inherent tension in this combination of views. On the one hand, the view that avowals are expressions of, rather than reports on, one's mental states encourages the view that they are not truth-evaluable. On the other hand, the view that avowals are self-ascriptions of mental states and are truth-evaluable encourages the view that they are not expressions of, but are rather reports on, one's mental states. Bar-On's attempt to show how these views can be held in combination involves arguing that the special epistemic security that avowals enjoy does not have an epistemic basis *at all*. This is the core of her account, and what makes it deflationary.

According to it, the special epistemic security that avowals enjoy has its source in the fact that subjects are immune to error in self-ascribing present mental states and (in the case of intentional avowals) in assignments of intentional objects to such states. This immunity to error is in turn due to the expressive nature of avowals, rather than to any recognition-based introspective self-knowledge. Thus, the asymmetry between certain self-ascriptions, on the one hand, and other self-ascriptions and other-ascriptions, on the other, is due to the fact that the former, but not the latter, are immune from error because they are expressions of rather than reports on one's mental states. This puts the account in the same category as those, mentioned above, which take some self-knowledge to be distinctive in being infallible but deny that there is any special method or means by which it is obtained.

Evidently, the main problem for all of the deflationary accounts considered here is that, although all concede that certain self-knowledge is authoritative in being infallible, infallibility does not account for authority in the sense that seems to matter for a genuine epistemic asymmetry between self-knowledge and knowledge of others. Further, there is a question whether, on deflationary accounts, subjects' infallible 'knowledge' of the contents of their own mental states deserves to be called knowledge at all.

43.3 Some Introspectionist Accounts

The preceding discussion encourages the thought that if self-knowledge is a genuine form of knowledge and there is an epistemic asymmetry between it and knowledge of the mental states of others, something like an introspectionist account is required to explain it. The Cartesian position is just one of these. In this final section I will explore this and a few other such positions and conclude by recommending my own.

As noted earlier, introspectionist views attempt to account for some or all of the features associated with certain self-knowledge by appeal to a special method or way in which one comes to know one's own mental states. Like deflationary accounts,

they recognize an asymmetry between self-knowledge and knowledge of others. Unlike deflationary ones, however, they attempt to explain how it is that at least some self-knowledge possesses some or all of the features associated with the asymmetry. It is important to the introspectionist view that the relevant knowledge concerns one's own current, conscious mental states, whether these are propositional attitudes or sensations.

While all introspectionist accounts appeal to a special method or way in which knowledge of one's own mental states is obtained, they differ in their accounts of what this method is, how it works, and whether the knowledge delivered by it has all of the features of self-knowledge listed at the outset of section 39.1 associated with the asymmetry between self-knowledge and knowledge of others. This section will discuss three such accounts. All are in agreement that one clear feature marks the asymmetry between self-knowledge and knowledge of others, namely the epistemic directness or immediacy of such knowledge, and that appeal to introspection is what explains this feature.

The first is what we earlier called the direct-observation account (see Gertler 2003). Its most famous expositor was Descartes, but a version of it was also held by Russell (1910), who maintained that subjects have direct epistemic access to their own mental states in virtue of being directly acquainted with them. More recently, versions have been endorsed by Chisholm (1981), Langsam (2002), and Chalmers (2003). While Chisholm's version is decidedly Cartesian, others, such as Chalmers's version, restrict the account to knowledge of one's sensation states. The reason why one might think it particularly plausible in the case of sensations, including visual experiences such as being appeared to redly, is that it is plausible to hold that no appearance/reality distinction applies to them.[13]

What distinguishes the direct-observation account from other introspectionist ones is not, or not merely, that it is observational, since, as we shall see, the inner-sense account to be discussed shortly is also an observational one. It is that the relation between the knowing state and the mental state known is held to be unmediated by any further mental state or mechanism in the subject. Chalmers (1996), for example, maintains that the knowing state is partly *constituted* by the phenomenal state known. This is thought to distinguish the kind of method involved in knowing one's own mental states from the kind of method involved in ordinary observation, or perception, of things in the world around one in at least two important respects. First, in ordinary perception the objects or phenomena perceived are distinct from, and can exist independently of, their being perceived. Second, and more importantly for present purposes, in ordinary perception there is a state of perceptual *experience* that mediates between objects of perception and beliefs/judgements about them, and a perceptual mechanism whose

[13] Chisholm gives as examples of 'self-presenting' properties not just feeling sad, but thinking about a golden mountain, being appeared redly to, and believing oneself to be wise, where a self-presenting property is defined as 'a property which is such that, if while having it, you consider your having it, then you will believe yourself to have it' (1981: 81). Further, 'the foundation of our knowledge consists of certain subjective—or 'Cartesian'—apprehensions'. (1981: 92).

operation is causally responsible for delivering that experience.[14] But according to the direct-observation account, knowledge of one's current conscious mental states is unmediated by any other mental states, experiential ones included. Such knowledge is epistemically direct and immediate in just the way described earlier in connection with the Cartesian model: it is non-evidence-based in being non-inferential and unmediated by any further mental state. One way of putting the point might be to say that knowledge of one's own current conscious mental states is constitutively rather than merely contingently and causally related to the first-order states that partly constitute such knowledge (see Shoemaker 1994).

This being so, one question that arises with this account, no less than with the deflationary ones discussed in Section 43.2, is whether on it self-knowledge has the status of being genuine knowledge. Wright (1989, 1998), for example, argues that knowledge involves the application of concepts, with respect to which error must be possible. If this is right, then in order to be able to know, say, that I am thinking that the apple on my desk is round, I need not just to think that the apple on my desk is round but also to have the concept of round and the concept of thinking and apply these correctly, where correctness implies the possibility of error. However, on the direct-observation account, where self-knowledge of mental states is construed in the way that Chalmers construes it, knowledge of one's own current conscious mental states is infallible. If I think that I am thinking that the apple in front of me is round, then I know that I am thinking that the apple in front of me is round. Wright's view is that on this kind of account self-knowledge is not genuine knowledge.

Even if Wright is mistaken about this and the account does not compromise the status of self-knowledge as knowledge, it does jettison the epistemic asymmetry between self-knowledge and knowledge of others, since this is connected with the notion of one's being epistemically better placed, or being better justified, than another to know one's own current conscious mental states. That notion seems to require that the reviewing thought and the thought reviewed be distinct existences, contingently and causally related. But this is just what the direct-observation account, specifically the constitution view, denies. Somewhat surprisingly, then, this account suffers from the same sorts of problems that bedevil deflationary accounts.

Further, as the discussion of deflationary accounts showed, there is a general problem reconciling authoritative self-knowledge with externalism about mental content, and the direct-observation account suffers from this problem as well. If externalism is true, what individuates the contents of certain mental states is the relations those states bear to factors beyond the bodies of their subjects—ones concerning which subjects themselves may be ignorant. So another's better knowledge of those relations may put her in a better position, or make her better justified than subjects themselves, to know what the contents of their current conscious thoughts are.

[14] Those who hold, with McDowell (1994), that perception itself is unmediated by an unconceptualized experiential state, but who hold that one's own perceptual states can be introspectively known in the observational way, will deny that there is this contrast but claim to be introspectionists with respect to self-knowledge (see Langsam 2002).

Later, at the end of this section, I will suggest a way of developing the direct-observation account that seems capable of avoiding these problems. But first let us consider a couple of other introspectionist accounts. According to one of these, the inner-sense account, introspection is viewed as a kind of inner perception. Versions of it have been held by Locke and Kant. Locke claimed:

> This Source of Ideas, every Man has wholly in himself . . . And though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense
>
> (1690/1975: II.i.4)

Kant described introspection as 'inner sense, by means of which the mind intuits itself or its inner state' (1787/1998: A23/B37; p. 157). More recently, it has been championed by philosophers such as Armstrong (1968, 1981) and Lycan (1987, 1996, 2004), both of whom also hold a particular view about what it is to be a conscious state (although commitment to this is not required by commitment to the inner-sense account of self-knowledge). According to this view, to be a conscious state is to be the object of higher-order awareness by a perception-like faculty. This view is known as the higher-order-perception theory of consciousness (one of a family of such theories, another of which is the higher-order-thought theory of consciousness—Rosenthal 1997).

The inner-sense account differs from the direct-observation one in important ways. First, and most obviously, it takes knowledge of one's own current, conscious states to be mediated by a causal, perception-like process involving a perception-like mechanism. Armstrong's position is that introspection involves a kind of self-scanning process by which subjects are made aware of their own mental states (1981: 61), involving an internal scanner or monitor that takes first-order states as inputs and outputs second-order ones, where these are distinct and contingently, causally related to one another. Lycan (1996: 14) describes it as 'the functioning of internal attention mechanisms directed at lower-order psychological states and events'. Armstrong's position takes the internal scanner to be in principle capable of scanning the lower-order mental states of others, so that the asymmetry between self-knowledge and knowledge of others is merely a contingent matter. Lycan's does not take the analogy with perception to be so complete as this, since it takes introspection to involve the use of primitive lexemes in a language of thought. This makes introspective self-knowledge ineffable.

Second, the inner-sense account attributes only one feature of those associated with the asymmetry between self-knowledge and knowledge of others described at the outset of our discussion to self-knowledge; namely, that it is direct and immediate, and this only in the sense that it is non-inferential. Self-knowledge has this feature because a subject need not know the causal relations that hold between her introspecting state and the state introspected when she knows her own mental state.

One objection to the account is that self-knowledge cannot be like perception because in perception there are three items (perceived object, perceptual experience,

and belief/judgement based on it) whereas in self-knowledge there are only two (first-order mental state and belief/judgement based on it) (Shoemaker 1994; Rosenthal 1997). To this the response has been made that the inner-sense view is not committed to construing introspection as like perception in every respect (Lycan 1996: 28).[15] Since, however, the presence of an experiential state seems to be essential to perception, the absence of any such state in introspective self-knowledge is a crucial difference.

Another objection is that self-knowledge cannot be like perception because in perception the relation between the object perceived and the perceptual state is contingent and causal. Applied to introspection this has the consequence that knowledge of one's current conscious states is fallible. The inner-sense account acknowledges both that it is possible for a subject to be in a first-order mental state and fail to know that she is because her internal monitor is malfunctioning, and that it is possible for a subject to be in a second-order, introspecting state without being in the corresponding first-order state (Lycan 1996). However, Shoemaker (1994) argues that if the connection between first-order and second-order states involved in introspection is contingent and causal, self-blindness (a condition in which one is introspectively 'blind' to one's own first-order mental states without any cognitive malfunction) should be possible, due to a breakdown in the internal monitor. But it is impossible, because normal rationality—specifically, rational-belief revision—requires access to one's own first-order states. So the requirements on normal rationality show that first-order states are not independent of second-order ones in the way required by the inner-sense account.

In response, one might agree that the inner-sense view is incompatible with Shoemaker's strong sense of 'rational' but claim that there is a weaker sense with which it is compatible. This is the sense that is relevant when one considers very young children, who are not yet capable of rational belief revision but are capable of adjusting their beliefs about the world in response to changes in their environments.[16]

But a more general worry about the account remains. Shoemaker's principal objection to the inner-sense account is that the relation that holds between the first-order and second-order states involved in self-knowledge is non-contingent and constitutive, and it is essential to human rationality that it be so. It is significant that the account marks an asymmetry between self-knowledge and knowledge of

[15] Thus, Lycan says: 'The inner-sense theorist does not contend (at least neither Armstrong nor I contend) that internal monitoring is like external perception in every single respect. And in particular, we should not expect internal monitoring to share the property of involving some presented sensory quality at its own level of operation' (1996: 28).

[16] In a similar vein, Lycan (1996) responds to Shoemaker's analogous argument with regard to pain by distinguishing between a strong and a weak sense of 'pain' and insisting, against Shoemaker's claim that

> at least certain kinds of pain behaviour . . . are intelligible as pain behaviour only on the assumption that the subject is aware of pain, for to see them as pain behaviour is to see them as motivated by such states of the creature as the belief that it is in pain, the desire to be rid of the pain, and the belief that such and such a course of behaviour will achieve that result
>
> (Shoemaker 1994: 274)

that the weak sense is a legitimate one.

others but takes it to consist merely in contingent facts about the method of knowing and the way in which the method operates. It fails to account for nearly all of the features associated with the asymmetry between self-knowledge and knowledge of others because it rejects the view that self-knowledge possesses them. But this robs the account of a satisfactory explanation of why at least some self-knowledge is epistemically special; a fact that the classically Cartesian account hoped to explain by appeal to introspection. Some will claim that this cannot be accounted for by any model that construes the relation between knowledge of one's own mental states and those states themselves as contingent and causal (Burge 1996, 1998; Gertler 2000).

Further, and relatedly, the inner-sense view suffers from a problem mentioned earlier in connection with deflationary positions, that of reconciling authoritative self-knowledge with externalism about mental content. According to it, the authoritative status of certain self-knowledge is a purely contingent matter, arising from the reliability of the introspective perception-like faculty. As it happens, this faculty is in general reliable. But despite this, if externalism is true, another may be better placed than I am to know the contents of my intentional states because she may have better knowledge of the beyond-the-body factors that individuate the contents of those states. This is because the internal scanner or monitor can only detect what is internal to the subject, not what is beyond the body.

A third type of introspectionist account agrees with the inner-sense view that knowledge of one's own mental states is perception-like, but denies that it involves the operation of any inner process or mechanism in addition to the process of perception itself. This view is known as the displaced-perception account (Dretske 1994, 1995, 1999). Whereas the inner-sense account takes introspection to be a matter of 'looking inwards', and attending to one's inner states, the displaced-perception account takes introspection to be a matter of 'looking outwards', and attending to the objects of normal perception, such items as trees, oranges, and human beings. Versions of it have been held by Evans (1982) and Dretske (1994, 1995, 1999, 2006). As Evans describes it:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me 'Do you think there is going to be a third world war?,' I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*.
>
> (1982: 225)[17]

This is *displaced* perception, since in perception one directs one's attention to the object perceived, whereas in self-knowledge one directs one's attention not to one's

[17] Evans continues:

> If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife.
>
> (1982: 225)

inner state, but outward, to a state or feature of the world beyond one. According to this account, one 'perceives' one's own mental state by perceiving something else.

Evans's version of the account speaks of self-ascriptions of beliefs, and his use of 'think' in the above example is most plausibly construed as functioning like 'believe'. The displaced-perception account seems ideally suited to explain cases like these. This is because it is plausible to maintain that it is in the nature of belief that beliefs aim at the truth of what is believed. So it is not surprising that when asked what I think about a possible or actual state of the world beyond me my attention is drawn not to the inner workings of my mind but to factors in that world. But things seem otherwise when the issue is not about what one believes, but about what one is currently consciously *thinking*. When asked what I am thinking, rather than what I think is or will be the case in the world beyond me, the more plausible view is that my attention turns not outward but inward. This is all the more apparent in cases of self-knowledge of one's sensation states.

Dretske's version of the account is connected with his teleorepresentationalist view of mental content. According to this, what gives a mental state its content—its meaning—is what its teleofunction is to indicate, what it represents. Just as what a petrol gauge's pointer's resting at E represents is that the petrol tank is empty, what my belief that water is transparent represents is that water is transparent. Dretske draws on analogies with devices such as petrol gauges and thermometers, where it is plausible to say that one becomes aware of one fact by becoming aware of another in general, in order to explain how it is that one becomes aware of one's own mental states of thinking, believing, and experiencing. So, just as one knows that the petrol tank is empty by perceiving the petrol gauge, one knows that one thinks that it is raining by perceiving the rain falling outside one's window. In general, one knows what properties one's own sensory or propositional state has not by attending to the state itself but by attending to what that state is of or about:

> To know what the experience is like, what properties it has, it is enough for the experiencer to 'look at' what the experience is an experience *of* (something that, as the experienc*er*, he cannot help but be doing). That will tell him what the relevant properties of the experience are.
>
> (Dretske 1999: 112)

Dretske's view is that the knowledge that one has of one's own mental states is special in that it is not only authoritative, but infallibly so.

This indicates that he takes at least some self-knowledge to be infallible. But in fact he goes on to say that more is involved in knowing whether I believe that *p* than just following whatever procedure it takes to determine whether *p*. At the very least it involves possessing the concept of belief, as it applies not just to oneself but also to others. Still, he claims, this further requirement need not involve looking inward. Note that Evans does not endorse this model for self-knowledge of experiences (1982: 226 ff.). Note also that, as in the case of at least some inner-sense models, the appeal to perception and 'looking inwards' in the displaced-perception account may be metaphorical (see n. 1). As the Evans quote in the text indicates, the key idea in the account seems to be that one comes to know what one thinks, believes, etc. not by appealing to any of one's inner states but by appealing to an outer state of the world, and this leaves it open whether the relevant state is one that is available to perception. Thanks here to John Williams.

But how is this possible? On the representational view of mental content the factors that determine the properties, or content, of an experience or other mental state are extrinsic, or relational. That is to say, the representational view is externalist about mental content. How, then, can one know authoritatively, let alone infallibly, what the properties of one's own propositional or sensory states are?

This is explained by appeal to the distinction between 'thing' awareness and 'fact' awareness, a distinction that is brought explicitly into play in order to help explain why the possession of these features by self-knowledge is compatible with externalism.[18] What I know authoritatively when I know that I am thinking that water is transparent, or that I am having an experience as of a red round apple, is *what* I am thinking or having, rather than *that* I am thinking or having it. That is, I am authoritative about the content of my thought or experience, but not about the fact that I am having a thought or experience with that content. As Dretske puts the point:

> Introspection is not how we know that we think and feel. It is how we know what we think and feel. Introspection is no more a way of knowing that we think and feel than is perception, our primary way of knowing what else is in the world, a way of knowing that there is something else in the world . . . We cannot see that there is an external world, although the things we come to know by seeing (that there is beer in the fridge, keys in my pocket) imply that there are things (namely, beer and keys) outside my mind.
>
> (1999: 137)

The idea is that when I am ϕ-ing that p, for some attitude ϕ and some propositional content p, I can infallibly know that it is p that I am ϕ-ing, but know only fallibly, if at all, and by some other means, that it is ϕ-ing (that p) that I am doing.

In order to see why Dretske thinks that self-knowledge of content is infallible, we need to understand an analogy he uses to demonstrate this. Consider an instrument, like a petrol gauge or a thermometer, that represents the value of a quantity, Q, say temperature. When it isn't malfunctioning, it correctly represents the quantity of a source to which it is connected. So, for example, when athermometer is functioning properly, when it registers (or 'says that') the source has Q of 37 °C, the source does have a Q of 37 °C. It carries this information, but does not say that it carries this information. But suppose we were to rewire it so that that one state, registering the source as having a Q of 37 °C, could 'say that' the source to which it is connected has a Q of 37 °C, but could also say something about itself—that it is representing

[18] Thus, he says:

> Everyone (even externalists) assume, mistakenly, that what we know by introspection is not only (in the case of thought) what we think, the *content* of our current propositional attitude, but also that we think it, the fact that we occupy a mental state having this proposition as its content. If this assumption is false, if what we know by introspection is that it is pumpkins one is thinking (wondering, worrying, deciding) about without knowing, at least not in the same way, that one is thinking (wondering, worrying, or deciding) about pumpkins, then there is no threat to externalism. What the teleosemanticist says is constituted by external, historical, relations—the fact that one is mentally representing pumpkins—is not the fact that introspection yields: that it is pumpkins one is mentally representing.
>
> (Dretske 2006: 76)

the source as having a Q of 37 °C. We might do this simply by affixing to it the label 'Value that Q is representing' (Dretske 1999: 134). This instrument now does two things by being in one state: it both represents the state of its source, and it represents itself as representing the state of its source. Although the instrument is fallible about the first thing it does (since it can malfunction), it is infallible about the second thing it does (even when it is malfunctioning).

One objection to this is that merely affixing a 'self-representing' label to a representational state is not sufficient to produce self-knowledge of that state (White 1987; Lycan 1996). The problem with it is that genuine self-knowledge involves not just representing the representational content of one's state to oneself, but representing that content to oneself *as* the representational content of one's state; and having a functional 'Value that Q is representing' does not suffice for this, since it does not suffice to ensure self-knowledge of the representational content of one's state. But even if it did, there would remain the problem that such knowledge, being infallible, is subject to the sorts of objections raised in connection with other introspectionist and deflationist accounts discussed above: that infallibility does not suffice to explain how a subject could be better placed than another to know the contents of her own state, and further, and more strongly, that it is questionable whether it deserves to be called 'knowledge' at all.

Because the account is externalist, it is also subject to the same sorts of objections based on switching examples discussed in Section 43.1. It is true that simple switching will not be a problem for the account, since the semantics associated with it is teleological, taking the determinants of mental content to be causal-historical, and a simple switch will not suffice for a change in the content of, say, a water thought. But the type of switching example discussed in Section 43.1 remains problematic for the account (Boghossian 1989). This is because, in the envisaged scenario, the subject has not only been unwittingly transported from earth to Twin Earth, but she also remains there for a long enough period for the causal-historical factors that determine content to determine her thoughts about the watery stuff to be water* thoughts. Evidently, apart from the inner-sense view, all of the accounts, both deflationary and introspectionist, thus far considered suffer from the same sorts of objections: one stemming from commitments to the infallibility of knowledge of one's own current conscious state, and the other stemming from considerations having to do with externalism about mental content. The inner-sense view does not suffer from objections of the former sort, but then again the asymmetry between self-knowledge and knowledge of others doesn't amount to much on this view either, since self-knowledge doesn't have the special epistemic status it is typically thought to have.

Let me conclude by recommending my own version of an introspectionist account, a version that I think is capable of avoiding many of these problems (see Macdonald 1995, 1998*a*, 1998*b*). Although it makes use of an observational analogy in order to defend a position on authoritative self-knowledge, it does not do so in what might appear to be the obvious way; namely, by appeal to something like an 'inner sense'. Rather, it appeals to more abstract and general features of

observation of external things, specifically features of observable properties, which help to explain our direct and immediate access to them, and this distinguishes the position from other introspectionist accounts discussed above. The position is principally intended to account for authoritative self-knowledge of intentional states of a particular kind—the current, conscious thinkings—but can be generalized to the more obvious case of sensations. The view is motivated by the thought that we can get a firmer grip on what makes for such knowledge by looking at how such notions as 'direct access' and 'immediate access' work in other areas where they have a natural home. One such place is in perception.

Consider certain observable properties of objects of perception, like being brown, or being rectangular. When I know that, say, the table visually present in front of me is brown, or that it is rectangular, this knowledge is in a certain sense direct and immediate. One explanation of how I can know directly or immediately that this instance is of a particular colour property, or of a particular shape property, is that it is presented to me *as* an instance of that property through my sense of sight. I simply see it *as* brown, *as* rectangular, non-inferentially and in an unmediated way.

This is not true of other properties of objects of perception. Water, for example, is an instance of the chemical structural property H_2O, but it is not manifested to me *as* an instance of that property through one of my senses. So certain properties seem to be ones of which we have direct or immediate awareness in perception because they are observable: whether objects are instances of them can be determined just by unaided observation (but not, perhaps, by unconceptualized experience) of those objects.[19]

Certain features of observable properties characterize their direct and immediate accessibility 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 that such properties are the ones by which objects that instance them are typically known in the first place. Knowing an object through instances of certain properties and not others favours certain ones epistemically. Another is that they *are* in general as they appear to normal perceivers in normal circumstances. 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.[20]

[19] The distinction I am after can be captured by means of the distinction between seeing *x* and seeing that *p*. I can both see that the table is brown and see the table's brownness; but although I can see that the liquid before me is H_20, I cannot see its 'H_20-ness'. Only the former counts as a case of direct epistemic access.

[20] However, unlike Wright (1988, 1992), who holds that one's authority with respect to one's own states consists in the fact that subjects' best opinions concerning those states fix the extensions of content types, the view I favour is that one's authority consists in the fact that one does not normally in

The rationale for focusing on the example of observable properties in perception is not to argue that self-knowledge is just like perceptual knowledge. There are clearly important and fundamental differences between these two sorts of knowledge. Crucially, there is a certain kind of phenomenology to perception of observable properties (for example, the experience of something's looking wet is distinctive and very different from the experience of something's looking red) that attaches to the contents of one's perceptual experiences. This seems to be lacking in the case of self-knowledge of one's thoughts, since, as noted earlier, there seems to be no analogue of a perceptual experience in the knowledge one has of one's own thoughts at all.

Despite this difference, there are important affinities that can help us to understand better the nature (and extent) of authoritative self-knowledge. In particular, the two above-mentioned features of observable properties of perception apply to contentful properties in these special cases of authoritative self-knowledge in a way that can help us to see why subjects have immediate awareness of them and so authoritative knowledge of their own thoughts. When I think about my own current, conscious intentional states while undergoing them, I think about them first and foremost *as* states of certain contentful types. Further, when I think about my states in this way, they are in general of the contentful types I take them to be. What makes for authority, when I have it, is that only I can be the subject of my intentional states, so that when I think about my intentional states as states of particular contentful types I am the only one to whom those contents appear to me in this epistemically basic and favoured way. So normally (i.e. barring special cognitive failures) when I think about my current conscious intentional states while undergoing them I am authoritative about the contents of those states.

Because the two features that characterize observable properties of perception are abstract and general, they are not tied to cases of observation alone. Those who appeal to such phenomena as 'intellectual experience' or 'intellectual intuition' in their accounts of authoritative self-knowledge may well appeal to such features (see, for instance, Burge 1996 and Bealer's theory of the a priori—1999). This distinguishes the position from other introspectionist ones. Further, the account is not tied to any particular conception of the *cogito*-type cases; specifically, to the conception that takes second-order thoughts about first-order ones to contain the first-order ones as constitutents. It can construe such cases as involving both a thinking of a first-order thought and a reflective thinking about that thought by the same subject, where the truth of the reflective one requires that it match the one reflected upon both with respect to content and with respect to the attitude. When successful, in reflection, the subject thinks a content of the same type as that •trkened in the first-order thought that is, she thinks the content •*again*) in the same attitudinal mode

• Q2
• Q3

reflection misidentify the object of one's reflection. Reflection is, in one respect at least, an appropriate characterization of the special relation that subjects' second-order thoughts bear to their first-order ones. In physical reflection—say, in a mirror—under certain ideal conditions, the object is not normally misrepresented. So the object is as it appears to be. But the reflection does not determine the object to be what it is. Similarly, in mental reflection the reflecting thought does not determine the thought reflected upon to be what it is.

and presents it *as* the content of the first-order thought.[21] This differentiates the view from the type of position held by Dretske (1999) discussed above, with its attendant difficulties. Thinking that redeploys thought content in that attitudinal mode brings a subject into direct epistemic contact with the content of the subject's first-order thought. This makes self-knowledge, even in the *cogito*-type cases, a genuine cognitive achievement.

Finally, the account seems to be compatible with externalism. The kind of epistemic access subjects have to the contents of their thoughts in such cases, in being direct and immediate, contrasts with the epistemic access others have to those contents. When I am currently consciously thinking and thinking about a thought with a given externalistically individuated content, another may be in a better position than I am to know what content is available for me to think, and so to think about. But that other is not in a better position to grasp, and so to know, the particular content that constitutes the subject matter of my thought, about which I am thinking. When *I* am both thinking it and thinking about it, I have, whereas another does not, a special kind of epistemic access to the content of that thought. Being in that position gives me an epistemic purchase on it that no other has. While it is true that this does not give me better discriminative knowledge of the contents of my thoughts, it is plausible to hold that, if externalism is indeed true, I *could* not be in a better position than another who has better knowledge of my environment to have such knowledge. But that does not compromise my authoritative knowledge of what I am currently consciously thinking (Burge 1988, 1998).[22]

References

Alston, W. (1971), 'Varieties of Privileged Access', *American Philosophical Quarterly*, 8: 223–41.

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

——— (1981), 'What Is Consciousness?', in Armstrong, *The Nature of Mind and Other Essays* (Ithaca, NY: Cornell University Press), 56–67.

Bar-On, D. (2004), *Speaking My Mind: Expression and Self-Knowledge* (Oxford: Oxford University Press).

Bealer, G. (1999), 'A Theory of the A Priori', *Philosophical Perspectives*, 13: 29–55.

Bermudez, J. (1998), *The Paradox of Self-Consciousness* (Cambridge, Mass.: MIT Press).

Boghossian, P. (1989), 'Content and Self-knowledge', *Philosophical Topics*, 17: 5–26.

[21] How might 'aboutness' or reference to the first-order content be secured? Roughly, in the way suggested by Davidson's paratactic analysis of sentences involving indirect discourse (1969/2001). Suppose I think to myself *I am currently thinking that water is transparent*. This second-order thought can be understood as a compression or abbreviation of something like: *I am currently thinking a thought whose content is the same as the following: water is transparent*. When I think this thought, I present my thought as having the same content as that of another current thought of mine, and I do this correctly if I am a *samethinker* with regard to these two thoughts (rather than a 'samesayer', as in Davidson's original suggestion).

[22] I would like to thank Bill Lycan, Graham Macdonald, and participants at the 2005 Australian Association of Philosophy Conference, Dunedin, New Zealand, for comments on this chapter.

Brueckner, A. (2001), 'Problems for the Agency Model of Self-knowledge', *Dialogue*, 40: 545–54.

Burge, T. (1979), 'Individualism and the Mental', *Midwest Studies in Philosophy*, 4: 73–121.

____ (1985), 'Cartesian Error and the Objectivity of Perception', in R. Grimm and D. Merrill (eds.), *Contents of Thought* (Tucson, Ariz.: University of Arizona Press), 62–76.

____ (1988), 'Individualism and Self-knowledge', *Journal of Philosophy*, 85: 649–63.

____ (1996), 'Our Entitlement to Self-knowledge', *Proceedings of the Aristotelian Society*, 96: 91–116.

____ (1998), 'Memory and Self-knowledge', in P. Ludlow and N. Martin (eds.), *Externalism and Self-Knowledge* (Stanford, Calif.: CSLI), 351–70.

Chalmers, D. (1996), *The Conscious Mind: In Search of a Fundamental Theory* (Oxford: Oxford University Press).

____ (2003), 'The Content and Epistemology of Phenomenal Belief', in Q. Smith and A. Jokic (eds.), *Consciousness: New Philosophical Perspectives* (Oxford: Oxford University Press), 220–72.

Chisholm, R. (1981), *The First Person* (Minneapolis, Minn.: University of Minnesota Press).

Davidson, D. (1969), On Saying That', *Synthese*, 19: 130–46; repr. in Davidson, *Inquiries Into Truth and Interpretation: Philosophical Essays*, 2nd edn.• (Oxford: Clarendon, 2001), 93–108.

• Q4

____ (1984), 'First Person Authority', *Dialectica*, 38: 101–11; repr. in Davidson, *Subjective, Intersubjective, Objective* (Oxford: Clarendon, 2001), 3–14.

____ (1987), 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 60: 441–58; repr. in Davidson, *Subjective, Intersubjective, Objective* (Oxford: Clarendon, 2001), 15–38.

____ (1988), 'The Myth of the Subjective', in M. Benedikt and R. Berger (eds.), *Bewusstsein, Sprache und die Kunst* (Vienna: Österreichischen Staatsdruckerei), repr. in M. Krausz (ed.), *Relativism: Interpretation and Confrontation* (Notre Dame, Ill.: University of Notre Dame Press, 1989), 159–72, and in Davidson, *Subjective, Intersubjective, Objective* (Oxford: Clarendon, 2001), 39–52.

____ (1989), 'What is Present to the Mind?', *Grazer Philosophische Studien*, 36: 3–18; repr. in Davidson, *Subjective, Intersubjective, Objective* (Oxford: Clarendon, 2001), 53–68.

____ (1991), 'Epistemology Externalized', *Dialectica*, 45: 2–3, 191–202; repr. in Davidson, *Subjective, Intersubjective, Objective* (Oxford: Clarendon, 2001), 193–204.

____ (2001), *Subjective, Intersubjective, Objective* (Oxford: Clarendon).

Descartes, R. (1641/1984), 'Meditations on First Philosophy', in *The Philosophical Writings of Descartes*, ii, trans. J. Cottingham, R. Stoothoff, and D. Murdoch (Cambridge: Cambridge University Press), 1–398.

Dretske, F. (1994), 'Introspection', *Proceedings of the Aristotelian Society*, 94: 263–78.

____ (1995), *Naturalizing the Mind* (Cambridge, Mass.: MIT Press).

____ (1999), 'The Mind's Awareness of Itself', *Philosophical Studies*, 95: 103–24.

____ (2006), 'Representation, Teleosemantics, and the Problem of Self-knowledge', in G. Macdonald and D. Papineau (eds.), *Teleosemantics* (Oxford: Oxford University Press), 69–84.

Evans, G. (1982), *The Varieties of Reference* (Oxford: Clarendon).

Gallois, A. (1996), *The World Without, The Mind Within: An Essay on First-Person Authority* (Cambridge: Cambridge University Press).

Gertler, B. (2000), 'The Mechanics of Self-knowledge', *Philosophical Topics*, 28: 125–46.

____ (2003), 'Self-knowledge', in E. Zalta (ed.), *Stanford Encyclopedia of Philosophy*, Spring 2003 edition, <http://plato.stanford.edu/archives/spr2003/entries/self-knowledge/>, accessed 2008.

Heil, J. (1988), 'Privileged Access', *Mind*, 97: 238–51.
____ (1992), *The Nature of True Minds* (Cambridge: Cambridge University Press).
Kant, I. (1787/1998), *The Cambridge Edition of the Works of Immanuel Kant: Critique of Pure Reason*, trans. and ed. P. Guyer and A. Wood (Cambridge: Cambridge University Press).
Langsam, H. (2002), 'Externalism, Self-knowledge, and Inner Observation', *Australasian Journal of Philosophy*, 80: 42–61.
Locke, J. (1690/1975), *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon).
Lycan, W. (1987), *Consciousness* (Cambridge, Mass.: MIT Press).
____ (1996), *Consciousness and Experience* (Cambridge, Mass.: MIT Press).
____ (2003), 'Dretske's Ways of Introspecting', in B. Gertler (ed.), *Privileged Access and First Person Authority* (Aldershot: Ashgate), 15–29.
____ (2004), 'The Superiority of HOP to HOT', in R. J. Gennaro (ed.), *Higher-order Theories of Consciousness* (Amsterdam/Philadelphia, Pa.: Benjamins), 93–113.
Macdonald, C. (1995), 'Externalism and First-person Authority', *Synthese*, 104: 99–122.
____ (1998*a*), 'Externalism and Authoritative Self-knowledge', in C. Wright, B. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 123–54.
____ (1998*b*), 'Self-knowledge and the 'Inner Eye'', *Philosophical Explorations*, 1: 83–106.
McDowell, J. (1986), 'Singular Thought and the Extent of Inner Space', in P. Pettit and J. McDowell (eds.), *Subject, Thought, and Context* (Oxford: Clarendon), 137–68; repr. in McDowell, *Meaning, Knowledge, and Reality* (Cambridge, Mass.: Harvard University Press, 1998), 228–59.
____ (1994), *Mind and World* (Cambridge, Mass.: Harvard University Press).
____ (1998*a*), 'Response to Crispin Wright', in C. Wright, B. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 47–62.
____ (1998*b*), *Meaning, Knowledge, and Reality* (Cambridge, Mass.: Harvard University Press).
McGeer, V. (1996), 'Is ''Self-knowledge'' an Empirical Problem? Renegotiating the Space of Philosophical Explanation', *Journal of Philosophy*, 92: 485–515.
McGinn, C. (1982), *The Character of Mind* (Oxford: Oxford University Press).
McLaughlin, B. (forthcoming), 'Self-knowledge', *Macmillan Encyclopedia of Philosophy*.
Moran, R. (1997), 'Self-knowledge: Discovery, Resolution, and Undoing', *European Journal of Philosophy*, 5: 141–61.
____ (2001), *Authority and Estrangement: An Essay on Self-Knowledge* (Princeton, NJ: Princeton University Press).
Putnam, H. (1975), 'The Meaning of ''meaning'' ', in Putnam, *Mind, Language, and Reality: Philosophical Papers*, ii (Cambridge: Cambridge University Press), 215–71.
Rosenthal, D. (1997), 'A Theory of Consciousness', in N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness: Philosophical Debates* (Cambridge, Mass.: MIT Press), 729–53.
Russell, B. (1910), 'Knowledge by Acquaintance and Knowledge by Description', *Proceedings of the Aristotelian Society*, 11: 108–28; repr. in Russell, *Mysticism and Logic* (London: Allen & Unwin, 1963), 152–67.
Ryle, G. (1949), *The Concept of Mind* (New York: Barnes & Noble).
Shoemaker, S. (1968), 'Self-reference and Self-awareness', *Journal of Philosophy*, 65: 555–67.
____ (1994), 'Self-knowledge and Inner Sense', *Philosophy and Phenomenological Research*, 54: 249–314.
Sosa, E. (2003), 'Consciousness and Self-knowledge', in B. Gertler (ed.), *Privileged Access and First Person Authority* (Aldershot: Ashgate), 253–62.

White, S. (1987), 'What is it Like to Be a Homunculus?', *Pacific Philosophical Quarterly*, 68: 148–74.

Wittgenstein, L. (1953), *Philosophical Investigations* (Oxford: Blackwell).

Wright, C. (1988), 'Moral Values, Projection and Secondary Qualities', *Proceedings of the Aristotelian Society*, suppl. vol. 62: 1–26.

____ (1989), 'Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and intention', *Journal of Philosophy*, 86: 622–34.

____ (1992), *Truth and Objectivity* (Cambridge, Mass.: Harvard University Press).

____ (1998), 'Self-knowledge: The Wittgensteinian Legacy', in C. Wright, B. Smith, and C. Macdonald (eds.), *Knowing Our Own Minds* (Oxford: Oxford University Press), 13–46.

Queries in Chapter 43

Q1. Please check and confirm the spelling as well as author correction here.
Q2. Author edit not clear. Pleas check
Q3. Opening parenthesis missing. Please confirm.
Q4. Please confirm whether this dot (.) for edn. is fine or not.