

In C.Wright, B. Smith, and C. Macdonald, eds., *Knowing Our Own Minds* (Oxford: Oxford University Press, 1998), pp. 123-154.

Externalism and Authoritative Self-Knowledge

Cynthia Macdonald

Externalism in the philosophy of mind has been thought by many to pose a serious threat to the claim that subjects are in general authoritative with regard to certain of their own intentional states.¹ In a series of papers, Tyler Burge (1985a, 1985b, 1988, 1996) has argued that the distinctive entitlement or right that subjects have to self-knowledge in certain cases is compatible with externalism, since that entitlement is environmentally neutral, neutral with respect to the issue of the individuation dependence of subjects' intentional states on factors beyond their bodies. His reason is that whereas externalism—the view that certain intentional states of persons are individuation-dependent on objects and/or phenomena external to their bodies—is a metaphysical thesis, authoritative self-knowledge is an epistemological matter. This being so, there is no reason to suppose that the two need conflict with one another.

I agree, but do not think that Burge's account of the nature and source of a subject's distinctive entitlement to self-knowledge establishes it. Burge claims that certain forms of self-knowledge (such as the *cogito* cases, which involve thoughts of the form *I am thinking that I am thinking that p*, for some propositional content *p*) are contextually self-verifying. That is, a subject's thinking such thoughts makes those thoughts true. But this does not, he thinks, in itself account for the peculiarly authoritative position that the subject has with regard to knowing what contentful thoughts she is thinking.² He acknowledges, that is, that a substantial epistemology is needed to account for first-person authority, even in cases like the *cogito* ones, where one is presently consciously thinking about a thought while thinking it, and where the first-order thought is a constituent of the second-order, or reflective, thought. His view is that such an epistemology can account for the peculiar entitlement that one has to knowledge of one's own intentional states, without being justificatory—without consisting of reasons or other factors that a subject is able explicitly to bring to bear on it. The distinctive entitlement is explicable simply in terms of a

¹ See e.g. Burge 1985a, 1985b, 1988; Davidson 1988; Heil 1988, 1992; and Boghossian 1989.

² This point is especially clear in Burge 1996. The account given here is based on this source.

subject's being in a favoured position with regard to her own intentional states. The question is: what is the nature of this favoured position, and what is the source of the authoritative status it confers on subjects' self-knowledge?

Burge's claim is that the source of a person's distinctive entitlement or right to self-knowledge lies, first, in the role that first-person thoughts or judgements play in critical thinking, and second, in the non-causal, *constitutive*, relation that holds between reflective thoughts and their subject-matter, or contents. Subjects reason critically, or reflectively, and in order to do so, they must be capable of thinking first-person thoughts knowledgeably. Since this is necessary for critical thinking, and subjects do reason critically, they must be entitled to, and must possess, knowledge of their own intentional states. What explains the *distinctiveness* of this entitlement is the fact that, in cases of self-knowledge, one is the *sole* subject of both one's first-order and one's second-order states. This gives the subject a single perspective on her own states, which forges a rational connection between reflective thoughts and the subject-matter of those thoughts. Since this is one perspective, *what* one is thinking and *what* is being thought about cannot regularly come apart; unlike cases of third-person knowledge of a subject's contentful states, there is no 'space' within which to distinguish the perspective of the thinker from the perspective of what is being thought about.

This single perspective affords an explanation of a subject's distinctive entitlement to self-knowledge only against the background of the role which first-person thoughts or beliefs play in critical thinking. This is so, not in the sense that one must be reasoning critically when one thinks first-person thoughts, but in the sense that one must have the capacity to think critically in order to be *entitled* to think these thoughts. One cannot reason critically unless one not only has reasons for one's beliefs but recognizes these *as* reasons. One uses one's knowledge of what counts as good reasons for thinking that *p* to 'guide' one's first-order reasoning, *as* reasons for thinking that *p*.³ Critical reasoning requires that one be capable of evaluating one's first-order thoughts and beliefs in the light of the norms that govern reasoning. One is incapable of doing this unless one is able not only to think about one's thoughts, but also to know what thoughts one thinks. So such reasoning requires self-knowledge.

However, one can agree with Burge that first-person thoughts play a distinctive role in critical reasoning, and that subjects could only engage in such reasoning if they were in general entitled to knowledge with regard to at least certain of their own intentional states, without accepting his explanation for this entitlement. Christopher Peacocke (1996) has argued that it is one thing to hold that a subject must be

³ This is Burge's (1996) terminology.

entitled to first-person knowledge of her own intentional states if she is to be capable of reasoning critically, another to hold that the *source* of such entitlement lies in the role that such knowledge plays in critical reasoning. His objection is that one can reason critically without making use of self-ascriptions, and so without thinking knowledgeably about one's thoughts. Furthermore, there seem to be cases of reflective thinking which involve self-ascriptions but are not *truth*-seeking, and so do not count as cases of critical reasoning: as cases of evaluating, justifying, criticizing and amending one's first-order beliefs or thoughts. In day-dreaming, for example, one may think reflectively while suspending judgement on the truth of one's thoughts. This does not show that Burge's claim that one's ability to reason critically requires that one be entitled to knowledgeable self-ascriptions is false. But it does strongly suggest that the role that they play in critical thinking is not part of the source of subjects' entitlement to self-knowledge.

Moran's (1994) distinction between theoretical-descriptive and prescriptive bases for psychological ascriptions brings this out nicely. He invites us to consider a person who, in the process of wondering about her current intentional states, asks herself the question, 'What do I think about X?'. This question can be interpreted in two ways: (a) as a theoretical-cum-descriptive one about an attitude antecedently held, or (b) 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. His point (or one of them) 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 criticize, 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, then Burge's claims about the role of first-person judgements/thoughts in critical reasoning do not establish that this is the source of subjects' entitlement to self-knowledge. It is plausible that there is a more primitive form of critical reasoning which does not require knowledgeable self-ascription (Peacocke 1996), and it is plausible that self-ascription plays a *descriptive* role in reflective thinking that either does not involve critical thinking or, if it does, does so in addition to the prescriptive role. Further, it may be that the descriptive, and not the prescriptive, role is the source of our entitlement to self-knowledge.

The aim of this chapter is to explore and defend this latter claim further by developing a quasi-observational account of authoritative self-

knowledge.⁴ I want to explore this not only because Burge's claims leave room for the development of such an account, but also because his explanation for our entitlement to self-knowledge does not clearly suffice to explain why authoritative self-knowledge is compatible with externalism.

Burge notes that there is an asymmetry between first- and third-person knowledge of a subject's intentional states, in that, in the former, but not the latter, a single perspective is involved, and that this single perspective forges a rational, non-contingent connection between reflective thought and thought reviewed.⁵ Suppose that there is this asymmetry. Then the question arises as to why we should think that this, and the rational connection it forges, confers authority on the subject and not on others. *What* is it about the perspectives possibly 'coming apart' that makes others *less* authoritative than me about the contents of my intentional states, about *what* is known? Why is it not the case that others, in having another perspective on my states, are better placed than me to know what I am currently thinking? It is not obvious that different perspectives must be less authoritative. This would seem to be particularly true if externalism is true. So, unless these questions can be given satisfactory answers, the difficulty in seeing how authoritative self-knowledge is compatible with externalism remains. Burge (1985*b*, 1988) thinks that it is the *immediacy* of the relation between subject and content that this unified perspective confers that makes for one's authority with regard to one's own intentional states. Indeed, it is generally thought that a person's authority with regard to the contents of certain of her intentional states (e.g. the *cogito*-like ones) consists at least in part in the *direct accessibility* to subjects of

⁴ It is *quasi*-observational for two reasons. The first is that there is a clear disanalogy between the observational model upon which the account is based and cases of authoritative self-knowledge with regard to the position of the subject. The second is that authoritative self-knowledge involves both causal and rationalistic elements.

⁵ It is unclear why it should be assumed that this is generally the case, or that it must be so: if I can take a different perspective on other subjects, or on what is visually present before me, or, as in my own reasoning, on the soundness of that reasoning, it seems possible that my perspective as thinker might 'come apart', even regularly, from my perspective as subject of what is thought about. It seems that for Burge this possibility is ruled out by the fact that, in the *cogito* cases, the first-order states are not only thought while being thought about, but are literally constituents of the second-order, reflective thoughts. The result is that there are not here two thoughts, whose relations to one another might be such that a subject's perspective on one might differ from her perspective on the other. However, this construal of the *cogito* cases is not the only possible one, as discussion of objections to the observational model in section II makes clear; nor are the *cogito* cases the only ones concerning which the question of authoritative self-knowledge arises, as Burge (1996) expressly concedes.

their own states, in the sense that knowledge of those states is typically not based on evidence.⁶ This in turn is thought to make for an asymmetry between first- and third-person knowledge of subjects' intentional states, since the latter is typically based on evidence. This asymmetry poses two prima-facie problems for externalism.

The first, which is not the primary concern here, is how to reconcile such non-evidence-based knowledge with the externalist view that the contents of certain intentional states are individuation-dependent on factors beyond person's bodies. The problem is that externalism seems to require what self-knowledge evidently rules out, namely, that knowledge of one's own intentional states be based on empirical evidence. The second problem, which is the primary concern here, is not how first-person *knowledge* can be seen to be compatible with externalism, but how *authoritative* knowledge can be reconciled with that position. The problem is that beliefs that are not based on evidence are not generally thought to be more reliable than ones that are. However, first-person authority seems to ground such authority precisely in the fact that it is not based on evidence.

So it is the peculiar epistemic *directness*, or immediacy, of the relation between a person and the contents of her own intentional states when she thinks about them, which Burge and others emphasize as crucial to the authoritative status of self-knowledge, that seems to pose problems for externalism. Why? If, for instance, first-person knowledge were like one's non-evidence-based knowledge of certain empirical facts external to one's body—say, knowledge that the table before one is white—then there would be no special problem for externalism to explain. For knowledge of this kind does not confer upon a subject an epistemic advantage over others. What makes the authoritative status of self-knowledge peculiar to it is that one's *right* or *entitlement* to it is *independent* of knowledge of the factors that individuate intentional content. This conflicts with the belief, which externalism apparently fosters, that others who have access to such factors are equally, and perhaps better, placed to know what the contents of a subject's intentional states are than the subject. That subjects should actually be *better* placed seems to be inexplicable on an externalist account. Burge's account of our distinctive entitlement to self-knowledge does not resolve this apparent conflict. In his view, a subject's distinctive

⁶ See Burge 1979, 1985a, 1985b; Davidson 1987; Heil 1988; and Wright 1989. Some, like Wright, emphasize the non-evidence-based character of such knowledge, whereas others, like Heil, emphasize the fact that such knowledge is not empirically evidence-based. Alston (1971) 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.

entitlement to self-knowledge has its source in the role such knowledge plays in critical reasoning and in the distinctive constitutive (non-causal) relation that holds between a subject's reflective thoughts and their subject-matter. This, he holds, is independent of knowledge of factors that individuate intentional content. But then a subject's entitlement to self-knowledge seems to conflict with the intuition that others who know more than the subject about the environmental factors that help to individuate intentional content are better placed to know whether, for example, in the presence of water, the thoughts she is thinking are indeed *water* thoughts. How is it that a subject's relation to the contents of her first-order states is such that an expert's better knowledge of certain contents does not compromise that subject's authority with respect to knowing that those states have those contents?

My claim is that careful development of a quasi-observational model can shed light on, and help to give an illuminating answer to, this question. At the same time, it can provide an account of the source of subjects' entitlement to self-knowledge that does not conflict with the role that first-person thoughts or judgements play in critical reasoning. This claim will be defended in two stages. First, I shall develop an account of a subject's direct epistemic access to the contents of her first-order intentional states that is broadly speaking observational in trading on important affinities between this access and subjects' epistemic access to certain observational properties of objects. This account explains the *source* of subjects' entitlement to self-knowledge, but not the *distinctiveness* of that entitlement. So, second, I shall mount an argument for first-person authority which exploits these affinities plus the further fact that subjects are the sole subjects of their own intentional states. It is this latter feature of self-knowledge that makes subjects' entitlement to self-knowledge *distinctive*, and so explains the asymmetry between first- and third-person knowledge of subjects' intentional states.

The plausibility of the resulting position depends crucially on how the model of observational properties upon which it is based is developed, and in this connection two objections to the position need to be addressed. The first is that, whereas the relation between observational properties and perceptual states of subjects is a causal one, the relation between first- and second-order intentional states in cases of authoritative self-knowledge is not. If so, then *no* observational model is appropriate for deployment in an account of authoritative self-knowledge. The second is that, whereas there are three items involved in observational models—observational property, perceptual experience, and conceptual states, such as beliefs, which those experiences ground—there are only two items involved in cases of self-knowledge—first-order intentional state and second-order reflective state. This makes for a structural disanalogy, so to speak, between the

two types of case, which threatens the applicability of any observational model to cases of self-knowledge.

The final section of this chapter addresses these objections, first by discussing the issue of the types of cases of self-knowledge to which the observational model is best construed as being applicable, and second by sketching two different ways in which the model might be developed so as to preserve structural similarity. Neither of these ways is entirely unproblematic; but then the aim is not so much to defend one particular way of developing the observational model as to defeat one line of objection to its applicability to cases of self-knowledge. Thus, the aim is to develop in these two stages an account of first-person authority for a certain class of intentional states, the *cogito*-like ones, that is capable of explaining why such authoritative self-knowledge is peculiarly authoritative, and why this is compatible with externalism. This is clearly a restricted class of states, but it is the focus of much of the recent discussion of authoritative self-knowledge, and it is arguably the most important class, since states like these are the most obvious cases of such knowledge. Moreover, the account is capable of being generalized to other cases of self-knowledge.⁷ For the sake of simplicity, the discussion will be restricted to thoughts whose specifications invoke terms for natural kinds, a class of thoughts concerning which many consider externalism to be most plausible.

I Stage One: The Observational Analogy and the Argument for First-Person Authority

Consider a situation 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 H₂O, 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. Burge's response to this apparent conflict is to say that not all cases of epistemic right or entitlement involve justification by reasons that subjects are capable of employing explicitly, and that cases of authoritative self-knowledge count as among these. The question, then, is where best to locate the source of such distinctive entitlement. The view that I shall defend is that authoritative self-knowledge consists in a combination of two factors. One is the *epistemic directness* of subjects' access to the contents of their first-order intentional states in such cases; the other is the fact that

⁷ For example, to the kinds of cases that Peacocke (1996) discusses: cases where a subject is presented with a perceptual experience which is a memory, and self-ascribes a belief that *p* on the basis of the

in these cases subjects are the *only* ones to whom these states are directly presented. The former can be developed by giving an account of direct epistemic access which is based on an observational model. The latter introduces the idea of a single subject and perspective that figures centrally in Burge's account of the *distinctive* entitlement which subjects have to knowledge of their own intentional states. On the whole, I shall concentrate on the former in the remaining parts of this chapter, since this is where I think that the source of subjects' entitlement to self-knowledge lies. However, the latter is necessary to forging a rational connection between reflective thought and thought reviewed, which is also required for authoritative self-knowledge (Burge 1996).

Let us turn, then, to the notion of direct epistemic access. 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.⁸ 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*.⁹ 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.

⁸ See his *Meditations*, especially Meditation II: 'Finally, I am the same who feels, that is to say, who perceives certain things, as by the organs of sense, since in truth I see light, I hear noise, I feel heat. But it will be said that these phenomena are false and that I am dreaming. Let it be so; still it is at least quite certain that it seems to me that I see light, that I hear noise and that I feel heat. That cannot be false' (Descartes 1969: 175). And later: 'What, then, I who seem to perceive this piece of wax so distinctly, do I not know myself, not only with much more truth and certainty, but also with much more distinctness and clearness? . . . And further, if the [notion or] perception of wax has seemed to me clearer and more distinct, not only after the sight or the touch, but also after many other causes have rendered it quite manifest to me, with how much more [evidence] and distinctness must it be said that I now know myself, since all the reasons which contribute to the knowledge of wax, or any other body whatever, are yet better proofs of the nature of my mind!' (ibid. 178).

⁹ See Alston 1971; Heil 1988, 1992; Burge 1985*b*, 1988; Davidson 1984, 1987, 1988; and Wright 1989. Some, like Wright, emphasize the non-evidence-based character of such knowledge; others, like Heil, emphasize the non-empirically evidence-based character of such knowledge. For present purposes it does not matter which construal of epistemic directness is at issue, since both cause problems for externalism.

For Descartes, the special status of such thoughts that derived from this immediate, non-evidence-based relation between subject and thought made for a kind of *transparency* of the thought reviewed to the reviewing subject, but not to others. It was this difference between a subject's access to the contents of her own intentional states and others' access to those contents that conferred an epistemic advantage on the subject in the sense of better placing her to know what the contents of those states are. But what exactly is it about this non-evidence-based relation that better places a subject in her own case? Why is it that another's better knowledge of the external factors that individuate certain contents (e.g. *water* ones) does not compromise this position?

I think that the relevant notion of epistemic directness, of being non-evidence-based, has two features, which can best be seen in the first instance in the case of observational properties, where they have a natural home. Consider properties other than mental 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. One explanation of how I can know directly that the table is an instance of this particular shape property, or 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; 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 property, H_2O , but this instance is not manifested to me as an instance of that property through any 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. 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 recognizing another instance of that property as of that property when presented with it on another occasion, and this requires that one 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—say, brown—it is not sufficient that one sees an instance of that property: one must see it as an instance of that property.

Turn now to sensation properties, such as the property, *pain*. These also seem to be ones to which we have direct epistemic access, in the sense that we typically know non-evidentially that we are in states that are instances of such properties. The general conception of direct

epistemic access appears to apply here too: when subjects are in pain, their pains are manifested to them as instances of the property, *pain*.

However, 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. And 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. The second, much more important one 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.¹⁰ 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 through observation 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 this, two features of observational properties characterize their epistemic directness in a way that I think is appropriate, within limits, to the characterization of mental properties in certain cases of self-knowledge (i.e. the *cogito*-like ones). The first feature is that observational properties 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 observational 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.

The second, crucial feature of observational properties is that they *are* in general as they appear when instances of them are presented to normal perceivers in normal circumstances. Again, this is not a point about the nature 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 H₂O, but this is

¹⁰ See e.g. Wright 1987, 1989, 1992.

compatible with water's instantiating certain observational properties that are as they appear to normal subjects in normal circumstances.

Both of these features apply to mental properties, with the limitation 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. However, I think that these features also apply to intentional properties.

Consider the first. 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 this state cannot be known by means of other properties (intentional or non-intentional); so it is not that I cannot think of this 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. Given this, all that is needed for authority is that, when I think of my state in this way, that state could not but be of the particular contentful type by which I think it.

Consider now the second feature: that such properties are in general as they appear to their subjects. I think that this feature also applies to intentional properties. The reason, to be developed more fully in the next section, is that the relation between them and normal subjects is in important respects like that between observational properties and normal perceivers. Given this, and given that subjects are the only ones to whom their contentful types appear in an epistemically basic way in the *cogito*-like cases of self-knowledge, 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 first-person authority consists of three main premisses:

- (1) A subject S typically thinks about her own intentional states as states of particular contentful types (i.e. such contents are epistemically basic to S).
- (2) S's intentional states are of the contentful types that they appear to be (i.e. such contents are knowable by S in an epistemically direct way).

(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 (i.e. barring special cognitive failures),

(4) S is authoritative with regard to the contents of her own intentional states.¹¹

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 access to my states. However, their access is evidence-based.¹² 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. The second is that when I know my states in this evidence-based way, there is no first-person authority for me. The third is that 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

¹¹ This version of the argument differs from that advanced in Macdonald 1995, where what appears here as premiss (2) appears as a conclusion derived from what here are premisses (1) and (3). The reason for the difference is that to construe (2) as following from (1) and (3) is to encourage the view that authoritative self-knowledge is due to the fact that, in reflection, a subject's attitudes toward her first-order states actually *determine* the extensions of those states. 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 favoured here is that one's authority consists in the fact that one cannot in reflection misidentify the object of one's reflection, and that this is so because the nature of the thought reflected upon determines the nature of the reflecting thought. Reflection is, in one respect at least, an appropriate characterization of the special relation which subjects' second-order thoughts bear to their first-order ones. In physical reflection—say, in a mirror—under certain ideal conditions, the object cannot be 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 nature of the thought reflected upon determines the nature of the reflection. If this is right, then (2) should appear as an independent premiss in the argument.

¹² By this I do not mean inference-based. See Alston 1971 and Burge 1996

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.

How is this reconcilable with externalism? Externalism tells us that the contents of our thoughts are individuation-dependent on factors in the environment. So it is possible that, on a given occasion, unbeknownst to me, I may be thinking a *twater* thought rather than a *water* thought. But if this is so, how can I have authoritative knowledge about the contents of my first-order thoughts?

I can have such knowledge because externalism is not a thesis about individual cases. If, on a single occasion, I am (unbeknownst to me) presented with *twater* in an environment which has regularly presented me with *water*, then my *water* thought will be false. However, it will still be a *water* rather than a *twater* thought, since stability in the environment with which I interact is an externalist requirement for concept formation. So, even when my first-order thought is false because I am being presented with *twater* rather than *water*, my second-order thought that I am thinking a *water* thought will be both correct and authoritative. Indeed, only if externalism is not a thesis about individual cases can I possibly have first-order thoughts that are *mistaken* with regard to substance concepts such as the concept *water*. But I do have such thoughts. And authority is not impugned.

Look now at the case in which my environment *does* regularly present me with *twater*, not *water*. In that case, my first-order thought will actually be a *twater* thought. But in that case, my second-order thought about that first-order thought will actually be a thought about a *twater* thought. So, once again, authority is not impugned.

The result is that the common-sense intuition that we have authoritative knowledge about our own intentional states is preserved *irrespective* of the truth of our first-order thoughts. Not only does externalism fail to threaten it, but it actually looks as if the truth of externalism is required for the preservation of the intuition.

Does this authority extend to knowing whether two token thoughts had at different times are thoughts of the same type? Not on the account given here. For this account is restricted to knowledge of states that I am currently consciously undergoing and considering *while* I am undergoing them. That is to say, it is restricted to *cogito*-like cases. And my knowledge that two token thoughts, had at different times, are thoughts of the same (or different) type(s), in involving memory, does not count as a *cogito*-like case. Further, since memory is involved, error is possible. So it may not be that knowledge of this kind is in fact authoritative.

II Stage Two: Developing the Account

The argument just given for first-person authority trades crucially on there being certain clear analogies between intentional properties, which are thought while being thought about by the same subject, and observational properties. Without this, the account of direct epistemic access which forms the basis of the distinctive authority which 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 is capable of explaining a subject's entitlement to self-knowledge, is that they *are* in general as they appear to normal subjects in optimal circumstances. But note again that this feature, and the role it plays in the account of direct epistemic access, does not establish, and so is insufficient for, the *distinctive* authority that subjects have over others with regard to knowledge of their own intentional states; so it does not explain the asymmetry between first- and third-person knowledge of a subject's intentional states. What establishes *that* is that subjects are the *only* ones to whom their intentional states appear in the epistemically direct way. I shall say more about this peculiar role that subjects play with regard to their own states in the closing paragraphs of this section. For now I put it to one side in order to develop the crucial component of the strategy, an account of observational properties that is both applicable to and explanatory of the epistemic relation which subjects bear to their own states when consciously entertaining or thinking them while thinking about them.

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 normal subjects in normal circumstances. It has seemed to many that the way in which this feature works to explain direct epistemic access in the observational case is very different from the way in which it works in cases of self-knowledge.¹³ One reason that has been given is that the relation between first- and second-order intentional states in the paradigmatic, *cogito*-like cases of authoritative self-knowledge is not, as it is in the paradigmatic case of observation, perception, a causal one.¹⁴ Another is that, unlike the case of perception, where there is a state of perceptual *experience* which mediates between objects of perception and beliefs/judgements about them, cases of self-knowledge are not mediated by experiential states.¹⁵

Consider the first claim. Is it true that in the *cogito*-like cases, the relation between first- and second-order, or reflective, thoughts is non-

¹³ See e.g. Boghossian 1989; Burge 1996; Shoemaker 1994; Peacocke 1992, 1996; and McDowell 1994.

¹⁴ See e.g. Shoemaker 1994 and Burge 1996.

¹⁵ See e.g. Peacocke 1992 and Ch. 3 above.

causal? Both Tyler Burge (1996) and Sydney Shoemaker (1994) have claimed so, albeit for different reasons.

Burge's reason has to do with the way he thinks *cogito*-like cases ought to be construed. In his opinion, the proper way to view them is as cases where 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.

Now it may be that, understood in this way, it is inappropriate to view *cogito*-like cases of authoritative self-knowledge as explicable along the lines of the observational model. But it does not follow that the observational model is inapplicable to all cases of authoritative self-knowledge. For one thing, Burge himself (1996) considers his *cogito*-like cases to form a very small subclass of the class of cases of self-knowledge which he intends his account of epistemic entitlement to cover, and this class will almost certainly contain cases where it is appropriate to view a subject's first- and second-order states as distinct and causally related. I am thinking here not only of cases of memory, where one can be said to 'redeploy' (Peacocke 1996) contents of an earlier experience, but also of the kinds of cases mentioned at the outset in connection with Moran's distinction between descriptive-theoretical and prescriptive roles of critical reasoning, where an intentional state is consciously present and one reflects upon it (as one does in wondering what one is presently thinking).

So, even if the observational model is inapplicable to *cogito*-like cases as Burge construes them, this would not in itself invalidate it for all cases of self-knowledge where the question of a subject's epistemic entitlement may arise. It is a moot point whether at least some of these cases are *cogito*-like ones. But there is a further question here as to whether Burge's own construal of the *cogito*-like cases is the only, or the best, way of construing them. He understands such cases to be contextually self-verifying because in them one's first-order thought is literally a constituent of one's second-order thought. However, this feature of such cases could be explained differently. It may be 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 the same. 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 distinct from the reflective one to which it is causally related.¹⁶

In short, then, I fail to see why the fact that not all cases of authoritative self-knowledge may be ones where it is plausible to speak of a causal relation between reflective thought and thought reviewed prohibits altogether the applicability of an observational model in an account of authoritative self-knowledge. Further, I think it plausible that some of these other cases may be *cogito*-like, even if they do not conform to Burge's construal of *cogito*-like cases. Finally, though, even the cases of contextually self-verifying thoughts seem to me to be capable of being construed in a way that is compatible with the claim that there is a causal relation between reflective thought and thought reviewed. If one's interest in accounting for subjects' distinctive entitlement to self-knowledge is, as Burge's is, motivated in part by the desire to provide a uniform general account of all such cases, then this would recommend a construal of the *cogito*-like cases that brings them more into line with others in recognizing the presence of a causal element and a first-order state distinct from the reflective one. In recognizing this, however, another problem with the observational model arises. The presence of a causal element suggests that we should be able to conceive of rational subjects in whom the quasi-perceptual faculty has broken down, and that in such cases subjects would think thoughts, have pains, and hear sounds, but not know that they are thinking, feeling, and hearing. But it is not clear that we can make sense of this.¹⁷

However, commitment to the observational model does not require that we can. One reason for thinking that we cannot trades on an ambiguity in 'know', in 'know that they are thinking, feeling, and hearing'. It is true that, in the sense in which to know that one is in pain, or hearing sounds, just is to be aware, or conscious, of them, we cannot

¹⁶ One might object here that our second-order, or reflective, mental states seem not to be independent of the states that they reflect upon, in that our self-monitoring seems intimately connected with the identity of what is monitored, changing and reconfiguring it as well as representing it. If so, then it would seem that the relation between first- and second-order intentional states in cases of self-knowledge is not like that between objects of perception in normal observation, and the quasi-observational account is wrong. My view is that there may be cases of self-knowledge that are like this, but not all such cases are, and that Moran's (1994) distinction between theoretical-descriptive and prescriptive bases for psychological ascriptions highlights this fact. The construal of the *cogito*-like cases offered in this paragraph is one which treats the first-order intentional states as independent of the second-order, reflective ones: states that form the theoretical-descriptive base for psychological ascriptions. In being independent of the second-order states, these states are like normal objects of observation.

¹⁷ 17 Akeel Bilgrami (Ch. 7 below) makes much of this point.

make sense of having pains, or hearing sounds, and not knowing that we are feeling or hearing. But this sense does not introduce a distinction between a first-order and a second-order state: here to be in pain just is to be aware, or conscious, of it. So the fact that we cannot make sense of subjects being in pain and not knowing that they are in pain does not threaten the quasi-perceptual model at issue, where there are two items involved in self-knowledge: a first-order state upon which one is reflecting and a second-order, reflective state.¹⁸

Moreover, in the sense relevant to the argument for authoritative self-knowledge based on the observational model, where two items are involved, a first-order state and a second-order, reflective state, the model is not committed to the claim that one can make sense of subjects thinking thoughts, having pains, or hearing sounds, but not knowing that they are thinking, feeling, and hearing. What it *is* committed to is that subjects in whom the quasi-perceptual faculty has broken down would not know *through quasi-perception* that they are thinking, feeling, and hearing. Such subjects would be, in Shoemaker's (1994) terms, 'self-blind': they would be in mental states that they are undergoing but to which they are introspectively blind, just as a person who is blind-sighted sees things that she is unaware of seeing and so is having perceptual experiences of which she is introspectively unaware.

Shoemaker claims that such a possibility is required by certain perceptual models of introspective self-knowledge, given that such models must presume the independence of, and so the contingency of the relation between first-order state and second-order, reflective state. This in turn is imposed by the requirement that such states be causally related. Crucial to both sorts of case is the possibility that the subject be able to conceive of those facts or phenomena to which she is blind, so that she is capable of learning of those facts or phenomena by means other than perception or introspection. What matters is that she is incapable of learning about them *perceptually*, through her perceptual experiences in the visual case, and *introspectively*, through her awareness of her first-order mental states in the case of self-knowledge.

Shoemaker himself does not think that subjects in whom the quasi-perceptual faculty has broken down would not know that they are thinking, feeling, or hearing. However, he does think it impossible for subjects to be self-blind with respect to either their sensations or their

¹⁸ Shoemaker (1994) makes this particularly clear in his discussion of self-blindness. Self-blindness with regard to pain, he argues, is not about being in a certain sensation state—say, a state of pain—and not feeling it: 'It must not be supposed that these creatures do not *feel* their pains. Pain is a feeling, and what they are self-blind to are, precisely, their feelings of pain' (1994: 274). Shoemaker proceeds to argue that self-blindness with respect to sensations is impossible, not because it would require one to be in pain and not feel it, but because it would require one to be in pain, with its characteristic phenomenological feel, without being introspectively aware of it.

intentional states, and he thinks that this is sufficient to discredit observational models of self-knowledge. Since his reasons for thinking this cut to the very heart of observational models of self-knowledge, they need to be considered.

Shoemaker mounts two different, but compatible lines of argument against observational models of self-knowledge: one for the sensations and another for intentional states. With regard to the former, he argues that in order to make sense of a subject being self-blind to them, one must suppose that she is self-blind to a whole host of other mental states, ones which would make intelligible, say, pain behaviour, such as that one finds the sensation unpleasant, and so has a desire for this reason to be rid of the pain. But a state of pain that played only the causal role that pain plays with regard to behaviour without playing the total causal role usually associated with pain—namely, the role it plays in causing beliefs and desires which in turn cause and make intelligible pain behaviour—simply would not be pain.

The case regarding sensations is less central to the issues that are of concern here than that regarding intentional states. Still, it is worth remarking on it at least to the following extent. Subjects in whom the quasi-perceptual faculty has broken down—that is, subjects who are self-blind—are meant to be otherwise cognitively normal. However, there are many creatures, human beings included, who are otherwise cognitively normal and who seem clearly capable of feeling, or hearing sounds, without knowing that they are feeling or hearing. Dogs, cats, chickens, and human infants count as such. Are we to suppose that such creatures do not feel or hear? Evidently so, since, as Shoemaker says,

Part of the causal role of pain consists in its being caused by certain kinds of things—bodily damage of various sorts—and its causing behaviors, such as wincing, grimacing, and moaning, that can be involuntary and do not have to be seen as motivated or 'rationalized' by beliefs and desires. Certainly a state could play this role without its subject having introspective awareness of it. And such a state would be, if not bad in and of itself, at least indicative of something bad, namely bodily damage. This it would share with pain. But it seems obvious that a state that played this causal role, but did not play any more of the standard causal role of pain than this, would not be pain. Indeed, it would not be a mental state at all.
(1994: 274)

However, this response begs the question against one who thinks that the relation between first-order states of feeling and hearing and

secondorder, reflective states is in fact contingent. That person will insist, plausibly on the face of it, that creatures who are incapable of forming beliefs and desires on the basis of their sensations may none the less undergo them. Nor will it do to respond that human infants are in fact capable of knowing that they are feeling and hearing because they are capable of becoming human adults, and human adults are capable of such knowledge. The fact that human infants are capable of becoming human adults does not make them human adults, and so does not imbue them with the capabilities of such beings.

So Shoemaker's case against a quasi-perceptual model *vis-à-vis* sensations is not decisive. But as I have said, the main target here is not the claim that a quasi-perceptual model will not work for sensations, but that it will not work for intentional states. What is the case for this?

Shoemaker's claim here is that self-blindness with regard to one's own intentional states is also impossible. His reason is that, in being in such states, and in having normal intelligence, conceptual capacity, and rationality, one *automatically* has second-order beliefs: 'if one has an available first-order belief, *and* has a certain degree of rationality, intelligence and conceptual capacity (here including having the concept of belief and the concept of oneself), then automatically one has the corresponding second-order belief' (1994: 288). A rational agent—that is to say, an agent who has normal intelligence, rationality, and conceptual capacity—will behave in ways that give the best possible evidence that she is self-aware, and so not self-blind. Such a person will be self-aware in that she has *normal awareness* of her own intentional states, where this sometimes involves, but *does not require*, reasoning sequentially from one premiss to another by a series of steps.

. . . in order to explain the behavior we take as showing that people have certain higher-order beliefs, beliefs about their first-order beliefs, we do not need to attribute to them anything beyond what is needed in order to give them first-order beliefs plus normal intelligence, rationality, and conceptual capacity. . . . in supposing that a creature is rational, what one is supposing is that it is such that its being in certain states tends to result in effects, behavior or other internal states, that are rationalized by those states. Sometimes this requires actually going through a process of reasoning in which one gets from one proposition to another by a series of steps, and where special reasoning skills are involved. But usually it does not require this. I see an apple and I reach for it. It is rational for me to do so, and this can be shown by presenting an argument, a bit of practical reasoning that is available to me, in terms of my desires and my beliefs about the nutritional and other properties

of apples. But I needn't actually go through any process of sequential reasoning in order for the beliefs and desires in question to explain and make rational my reaching for the apple. And no more does the rational agent need to go through a process of sequential reasoning in order for her first-order belief that P, plus her other first-order beliefs and desires, to explain and rationalize the behavior that manifests the second-order belief that she believes that P. (Shoemaker 1994: 284-5)

As this passage makes clear, Shoemaker's view is that being self-aware requires no more than being in states of belief and desire that tend to cause and rationalize other intentional states and behaviour.

But what kind of self-awareness is this? The crucial phrase in the quoted passage is 'rationalize the behavior'. Shoemaker's self-aware person is one whose beliefs and desires rationalize her behaviour, not from her own perspective, but from the perspective of *others*. It is true that she is typically *capable* of reasoning sequentially, and so of rationalizing her behaviour from her own point of view, but such reasoning and rationalizing is not required for self-awareness. So self-awareness is something subjects can possess, not by virtue of exercising a first-person perspective on their own states, but simply by behaving in such a way as to be subjects of *others'* rationalizing strategies. But then I think that Shoemaker's use of the term 'self-aware' to mean 'introspectively self-aware' is misplaced: subjects whose rationalizing behaviour counts as such only from the perspective of others do not provide the best possible evidence for being introspectively aware of their own beliefs and desires. Conscious, yes, but introspectively conscious, no. Further, any conception of self-awareness stronger than this—say, in requiring sequential reasoning—would require reflecting on one's reasons *as reasons*, and so require a more robust account of introspective self-knowledge than Shoemaker is prepared to endorse.

Shoemaker's argument is effectively that one's first-order states do not in general provide *grounds* or *reasons* for one's having second-order beliefs. As he puts it, 'believing that one believes that P can be just believing that P plus having a certain level of rationality, intelligence and so on' (1994: 289). This view, which Peacocke (Ch. 3 above) calls the 'no reasons' view, is often contrasted with a view of self-knowledge which takes such knowledge to be based on a kind of inner observation. Shoemaker's argument is that self-blindness in the case of intentional states is impossible, but not, as in the case of sensations, because it is impossible that a subject should be in first-order intentional states and have beliefs about such states based on grounds *other* than introspective ones (say, behaviour, or the subject's awareness of her

own neurophysiological states). Rather, it is impossible because self-knowledge of one's own intentional states is not based on grounds or reasons of any kind. So the reason why self-blindness is impossible in this case is that the connection between having first-order beliefs and being able to self-ascribe them is too *direct* for there to be either some kind of inner awareness of or blindness to one's first-order states. Such knowledge is *baseless*.

It would not be inappropriate to respond to Shoemaker's argument here by pointing out that, as he has constructed the case of self-blindness, one might know of one's own first-order intentional states not by means of introspection, but by other perceptual means such as behaviour and/or awareness of one's neurophysiological states. And these means would seem to provide one with grounds or reasons for thinking that one is in certain first-order intentional states, albeit less direct grounds than one might have if one were to know of one's first-order states by inner awareness of them. Of course, *if* Shoemaker is right in thinking that such second-order knowledge is not acquired on the basis of any grounds or reasons, then this kind of response is irrelevant. But the question is whether he *is* right, and it would be question-begging to reject the possibility of self-blindness on this basis. More importantly, however, even supposing that he is right with regard to certain cases of self-knowledge, why should we think that all cases of self-knowledge are of this kind? I think that Shoemaker's reasons for thinking that they are turn out, on examination, to trade on an ambiguity in the term 'rationality' and its near cousin, 'self-aware'.

It has been said that it is essential to belief possession, and to the concept of belief, that beliefs are subject to rationalistic constraints. Thus to have a belief that p , for some propositional content p , is (amongst other things) to be disposed to assert that p , to engage in reasonings using the content that p , and to be prepared to revise the belief that p in the light of conflicting evidence.

In one, very minimal sense of the term 'self-aware', to be self-aware is simply to employ one's beliefs in rational behaviour, to have one's first-order beliefs cause one's other intentional states and behaviour in ways that can be rationalized. It is in this sense that one can attribute to children, who may not yet be capable of critically evaluating their reasonings to see whether, for example, they conform to *modus ponens*, self-awareness of their own states. It seems clear that this is the sense of self-awareness that Shoemaker has in mind when he claims that a subject who is self-aware is rational. Suppose he is right in thinking that engaging in a process of sequential reasoning to rationalize one's own behaviour is not required for rationality in general, and so is not required for self-awareness (in his sense of the term). Still, it does not establish what Shoemaker thinks it does with regard to the observational model, for two reasons.

First, the applicability of the observational model to introspective self-knowledge does not require that it fit every case of such knowledge. Nor is it plausible to think that it should. The concern here is with a certain class of intentional states: namely, those that one is currently, consciously thinking about while thinking them. It is these states concerning which the observational model seems most appropriate. But these are not the states with which Shoemaker is primarily concerned, if at all. So even if his claims regarding self-awareness and rationality are correct with respect to beliefs about one's beliefs in cases where one is *not* currently, consciously having them while considering them, nothing of interest would follow about the applicability of the observational model in cases where one *is* currently, consciously having such states while considering them.

But second, and more importantly, it may be that one can only engage in rational *deliberation* and *reflective reasoning*, reasoning about one's reasoning, if one is capable of thinking certain kinds of thoughts, thoughts of the form *I am thinking that I am thinking that p*, and so of being self-aware in a much stronger sense than Shoemaker is prepared to countenance. The reason is that, in order to deliberate, or to reason reflectively, one needs to be able to *examine* one's reasons, and to consider them as *reasons* for behaving one way or another, or for accepting or rejecting or revising one's other attitudes. To do this evidently requires being able to view them as *objects* of reflection or deliberation. As Burge puts the point,

To be capable of critical reasoning, and to be subject to certain rational norms necessarily associated with such reasoning, some mental acts and states must be *knowledgeably* reviewable. . . . Critical reasoning is reasoning that involves an ability to recognize and effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by an appreciation, use, and assessment of reasons and reasoning as such. (1996: 98)

Burge considers it essential to the ability to reason critically that one be capable of thinking *cogito*-like thoughts, and he associates this ability with being rational in some sense of 'rational'. This sense of 'rational' is clearly stronger than what Shoemaker has in mind. Now it may be that one can be rational without being a critical reasoner (although it is less clear whether one can be rational without being *capable* of being a critical reasoner). So it may be that one can be rational in Shoemaker's weaker sense without being rational in the stronger sense associated with the ability to reason critically (as Burge construes it). However, if, as seems clear, it is possible to be rational in this stronger sense, then it is also possible to be self-aware in a sense

stronger than that associated with Shoemaker's conception of rationality. It is to be capable of reflective reasoning, of reasoning about one's reasoning, where the so-called objects of such reasoning are one's own first-order intentional states. And if, as Burge argues, this requires the ability to think *cogito*-like thoughts, thoughts about what one is currently, consciously thinking, where one is both thinking and thinking about a propositional content, then not all self-knowledge can be construed in the way that Shoemaker construes it.

Shoemaker in fact recognizes this, but points out that his conception is the one that is in fact met by most people most of the time, and that the notion of rationality that involves engaging in critical reasoning is an ideal that most of us regularly fail to meet. However, that an ideal regularly fails to be met does not thereby make it *unapproachable*, and this is all that is necessary for the ability to engage in critical reasoning to be serviceable as a requirement on rationality in general. Further, that there is this stronger sense of rationality opens up conceptual space for an account of introspective self-knowledge for a certain class of cases, the *cogito*-like ones, which is broadly observational. This should be of interest to anyone whose concern with self-knowledge is a concern with *authoritative* self-knowledge. The *cogito*-like cases are central, not peripheral, to the issues here, precisely because they count as paradigmatic ones of authoritative self-knowledge. And the observational model is of use in helping to make intelligible these cases of self-knowledge precisely because it promises to deliver an explanation of the special epistemic right, or entitlement, that subjects have with regard to knowledge of the contents of certain of their own intentional states in just these cases.

However, the presence of a causal element in the observational model is not the only, nor the most serious, obstacle to basing an account of authoritative self-knowledge on such a model. What really seems to bother those who firmly reject this model is the thought that, to make it work, one would need to suppose that there is, in cases of self-knowledge, an analogue of *perceptual experience*, a phenomenal awareness of one's intentional states, which mediates between the 'object' of reflective awareness (the first-order intentional state) and the state of reflective awareness itself. Peacocke, for example, rejects observational models on the grounds that in perception there is a state of perceptual experience which is partly non-conceptual and whose non-conceptual component figures in the possession conditions for observational concepts, thereby mediating between beliefs and judgements that employ such concepts and the objects that they are judgements about. On his view, because there is no such analogous non-conceptual component in cases of self-knowledge, observational

models (specifically perceptual ones) are not applicable to such cases.¹⁹

The complaint here is that observational models won't work because, whereas in observation there are three items—object, perceptual experience, and belief/judgement which experience grounds in the sense of providing 'primitively compelling grounds' (causes which are also reasons) for judgements about the perceptible properties of objects in subjects' environments—in self-knowledge there are only two items—'object' (first-order intentional state) and belief/judgement about that state.

Whether this complaint really bites, however, depends, first, on how one views the observational model as operating in perception itself, independently of the issue of its applicability to cases of self-knowledge, and second, on how it is meant to extend to cases of self-knowledge. The thesis argued for in the previous section was that certain features of observational properties characterize their epistemic directness in a way that is applicable to mental properties in certain cases of self-knowledge. The question is: what items do these features relate to in the observational case, and can these features be said to be applicable in an analogous way to cases of self-knowledge? That there are (if there are) three items in the former case and only two in the latter is not in itself decisive against the observational model if, in the perceptual case, the features of epistemic basicness and epistemic directness relate only two of those three items. Alternatively, if there are only two items involved in the perceptual case—if, for example, there is no state of non-conceptual perceptual experience that mediates between conceptualized contents that figure in beliefs and the objects that they are about—then again, for a different reason, the objection is misplaced.

This suggests two rather different ways in which to develop the observational model for the purpose of extending it to cases of self-knowledge. One way would be to hold that in perception itself there is no state of perceptual experience which is even partly non-conceptual. In effect, this would be to deny that in perception there are three items. A version of this view has recently been advanced by John McDowell (1994), who has argued that a satisfactory naturalistic theory of mind can succeed only if the 'space' of reasons extends outward, into perception itself, so that there is no epistemic 'gap', so to speak, between mind and world. For McDowell, there is reality, and there are states of mind which, in perceptual experience of the world, deploy concepts that act both as passive receptors of information from the world (are 'passively drawn into play in experience') and as active in playing a role in judgement and, more generally, in critical reasoning. Whatever the merits or demerits of this view, the perceptual model with

¹⁹ See e.g. Ch. 3 above.

which it works, in dealing with two items rather than three, is structurally similar to that of self-knowledge. Might this model provide a basis for an account of authoritative self-knowledge that deploys the features of direct epistemic access, namely, epistemic basicness and epistemic directness?

McDowell claims that his view has no need for a notion such as that of direct epistemic access. This is because, on his view, the 'space' of reasons extends into the world itself: 'The idea of conceptually structured operations of receptivity puts us in a position to speak of experience as openness to the layout of reality. Experience enables the layout of reality itself to exert a rational influence on what a subject thinks' (1994: 26). What he seems to mean by this is that rational connections hold, not only between intentional and other states of mind, but between them and the world beyond the mind, and this because 'there is no ontological gap between the sort of thing one can mean, or think, and the sort of thing that is the case. When one thinks truly, what one thinks *is* the case . . . so there is no gap between thought and the world' (1994: 27). Although it is not entirely clear what exactly McDowell's thesis is here, it seems to be that in so far as the world, in being thinkable, is a conceptualized world, it exerts a rational influence on experience and judgements.

The existence of rational connections between mind and world sets McDowell's perceptual model apart from the standard conception of an observational model. If it were to work—if there really is no ontological gap between thought and the world—there would be no need for the kind of observational model envisaged here, no explanatory work for the features associated with direct epistemic access to do. For if there is no 'boundary' between thought and world—if they make direct contact—then, analogously, so do first- and second-order thoughts in cases of self-knowledge. Nor would there be any need for an additional component in the argument for first-person authority to forge rational connections between thought reviewed and reflective thought. Since rational as well as causal connections are built into the perceptual model itself, its application to cases of self-knowledge seems to be a relatively straightforward matter.

However, I am not convinced that the kind of account of the relation between mind and world that McDowell envisages does work, hence that there is no need for an observational model that deploys features associated with direct epistemic access. This is because I cannot see how McDowell can make intelligible the idea that there is 'friction' between mind and world, which he recognizes to be critical to fending off the charge of idealism, *without* supposing there to be an 'ontological gap' between mind and world, and so without supposing there to be a boundary between the space of reasons and the world. So long as there is this gap, there is a need to bridge it; and the features associated with direct epistemic access are meant to do just that.

My point is not that McDowell's two-item perceptual model cannot be put to use in mounting an argument for first-person authority based on what I conceive of as an observational model. I think it can. But then one would have to abandon McDowell's position on the relation between thought and the world. But since this claim is itself problematic, abandoning it, while retaining the model, is a promising strategy for extending the observational model to cases of self-knowledge.

However, this strategy will not be viewed as a way of carrying the observational model forward by those who think that there *is* in perception a state of perceptual experience which is partly non-conceptual and which mediates between objects of perception and beliefs and judgements which perceptual experience grounds. Such a model of perception has recently been elaborated upon and defended by Peacocke (1992). Peacocke is concerned with a much larger project of providing a theory of concepts as epistemic capacities. One aspect of this is to provide non-circular necessary and sufficient conditions for the possession of concepts of various sorts, observational (perceptual) concepts being a central case. His claim is that such accounts will not be circular if mention of the concept whose possession condition is in question does not occur within the scope of psychological attributions to the possessor. So, for example, this is his account of the possession conditions for the concept *red*:

The concept *red* is that concept *C* to possess which a thinker must meet these conditions:

1. He must be disposed to believe a content that consists of a singular perceptual-demonstrative mode of presentation *m* in predicational combination with *C* when the perceptual experience that makes *m* available presents its object in a 'red' region of the subject's visual field and does so in conditions he takes to be normal, and when in addition he takes his perceptual mechanisms to be working properly. The thinker must also be disposed to form the belief for the reason that the object is so presented.
2. The thinker must be disposed to believe a content consisting of any singular mode of presentation *k* not meeting all the conditions on *m* in (1) when he takes its object to have the primary quality ground (if any) of the disposition of objects to cause experiences of the sort mentioned in (1). (Peacocke 1992: 7-8)

Peacocke notes that similar non-circular possession conditions can be formulated for concepts in terms acceptable to those who reject sensation properties (and so reject states with non-conceptual perceptual content). However, his view is that there are such states, that these provide 'primitively compelling' grounds for beliefs and judgements employing observational concepts about the perceptible properties of objects in perceivers' environments, and that these grounds are not only causes, but reasons for which subjects judge and act as they do.²⁰ The view is in fact much richer and more sophisticated than I can do justice to here. But two points are noteworthy. First, for Peacocke perceptual experiences have two layers of non-conceptual content, the second of which is proto-propositional, and both of which require mention in possession conditions for observational concepts such as *red*. Protopropositional components include individuals, properties, and relations, and are necessary to capture subjects' perceptual experiences of such relations as square, curved, symmetrical about, parallel to, and so on. Second, although perceptual experiences have a non-conceptual component which figures in the possession conditions for observational concepts such as *red*, typically for a thinker who does possess such a concept, that concept will enter into her perceptual experiences:

If we take an observational concept we do in fact possess, then certainly we will have difficulty in successfully imagining an experience in which the observational property is perceived but not conceptualized as falling under that observational concept. This is not surprising. For those who possess an observational concept, it enters the representational content of their experiences, and derivatively, of their imaginings. (1992: 81)

Suppose it is true that perceptual experiences have a non-conceptual component which mediates between concepts employed in beliefs and judgements about objects of perception and the objects themselves. Would this make the observational model envisaged here inapplicable to cases of self-knowledge on the assumption that there is no analogous mediating experiential state in such cases?

That depends partly on whether the features of the observational model that are meant to apply to cases of self-knowledge hold in the

²⁰ A primitively compelling ground for the possession of a given concept *C* is a ground (1) that the subject finds compelling, but (2) not because she has inferred it from other principles and/or premisses, and (3) whose correctness the subject need not take as answerable to anything else in order to possess *C*. See Peacocke 1992: 6.

perceptual case between all three items. But it is implausible to suppose that the features associated with direct epistemic access do so. On Peacocke's model, perceptual experience both mediates between perceptual belief and objects of perception and provides primitively compelling grounds for judgements about the perceptible properties of objects in subjects' environments. Given this, the plausible line to take is that the features associated with direct epistemic access hold in the first instance between only two of the three items involved in perception, the perceptual experience and the beliefs and judgements which that experience grounds.

This removes one obstacle to extending the observational model to cases of self-knowledge, the response to it being that although there are indeed three items in perception and two in self-knowledge, the epistemic features in play in perception of observational properties that are relevant to the applicability of the observational model to cases of self-knowledge hold between only two of the three. This makes the cases of perception and self-knowledge structurally analogous. And this is so irrespective of the fact that since the so-called object of perception drops out as irrelevant in cases of self-knowledge, judgements in these cases are about the very first-order states to which the reflective states bear relations of epistemic directness. Nor is the difference between the two cases here as great as it may seem to be at first. For, in cases of hallucination, there is no object of perception about which to judge; and in reflective thinking, thoughts are directed beyond one's thoughts, toward the world, and often ground judgements about it.

But now there is another objection to consider. Peacocke's model is one in which, prior to concept possession, only the non-conceptual component of perceptual experience is one to which a subject is related. Further, it is this, and not any conceptual component that may subsequently enter into perceptual experience, that provides subjects with primitively compelling grounds for forming beliefs and judgements employing perceptual concepts about the perceptible properties of objects in their environment. Given this, the objection is that the features of direct epistemic access apply, if at all, to the non-conceptual component of such experiences, but that there is no such component in cases of self-knowledge, so the observational model cannot be extended to such cases.

I think that the right way to respond to this is as follows. It may be that, prior to concept possession, the perceptual experiences to which subjects are related contain no conceptual component, and further, that such experiences provide primitively compelling grounds for forming beliefs containing perceptual concepts. However, the cases of interest here are not those of subjects who lack observational concepts, but rather those of subjects who, in having self-knowledge, are in full possession of these and many other concepts, notably the concept of belief. These cases are ones where subjects' perceptual experiences

will generally speaking contain observational concepts. Further, in them it is plausible to say that conceptual and non-conceptual components are inseparable from the perspective of the subject. That there may be perceptual experiences in which only non-conceptual components come into play does not establish that, in experiences in which both non-conceptual and conceptual components figure, they are separable. But if so, then the difference between the perceptual case and that of self-knowledge is a difference of degree rather than kind. For if the conceptual and non-conceptual components of perceptual experience are inseparable in experience for subjects who possess perceptual concepts, then subjects cannot be said to have direct epistemic access to one *rather than* another component of perceptual experience. (This marks a difference between Peacocke's notion of primitively compelling grounds and the notion of direct epistemic access, since the latter, on the account sketched earlier, is intentional, whereas the former is not.)

There is a further point to be made here. Peacocke supposes perceptual experience to contain a non-conceptual component. But what about the perceptual beliefs which those experiences ground? Do they too contain such a component? If so, what about beliefs about those beliefs? And if not, why not? On what basis are we to judge the point at which non-conceptual components no longer enter into the contents of intentional states? It seems unlikely that this can be determined on entirely a priori grounds. But then the applicability of the observational model to cases of self-knowledge cannot be ruled out on a priori grounds either.

Is there any need to appeal to the features of direct epistemic access in the kind of perceptual model that Peacocke employs? I have suggested one difference between Peacocke's notion of primitively compelling grounds and the notion of direct epistemic access in play in the observational model envisaged here: the latter is intentional, whereas the former is not. For one to have direct epistemic access to a colour property—say, brown—it is not sufficient that one sees an instance of that property: one must see it as an instance of that property. This difference might be exploited in accounting for a thesis which Peacocke takes to be central to his view and to his account of our entitlement to self-knowledge: namely, that the non-conceptual content of perceptual experiences can yield 'not merely reasons but good reasons' for beliefs formed on the basis of them. It has been said that Peacocke's defence of this thesis falls short of what is needed, since it appeals to reasons *for which* a subject forms a belief, whereas what is needed is a *subject's* reasons. Reasons for which a subject forms a belief may exist even when the subject has no reasons. Reasons are goal-oriented, and they need to be so oriented from the perspective of the subject herself.²¹

²¹ See McDowell 1994: pt. 2.

This opens up explanatory space for the notion of direct epistemic access to do some useful work. It may be that the non-conceptual component of perceptual experience provides primitively compelling grounds for forming beliefs employing perceptual concepts. But these may be too primitive to provide reasons, in the sense of a subject's reasons, as well as causes. Conceptualized perceptual content may be what is required, and this is where the notion of direct epistemic access might do some real explanatory work.

What I have attempted to make plausible here is the idea that the *source* of subjects' authority with regard to certain of their first-order states is other than the role that first-person judgements or self-ascriptions play in critical reasoning. If the argument goes through, the source lies in the epistemic relation which subjects bear to the contents of their first-order states, and the features of contentful properties. But the observational model upon which the argument is based can do no more than explain how it is possible that subjects should have direct epistemic access to such properties; it cannot account for the peculiarly authoritative status that subjects have with regard to their own intentional states.

As I have indicated, the additional factor necessary to account for first-person authority is that, in cases of self-knowledge, the subject is the *only* one to whom her first-order intentional states appear directly. On this matter I have little to add to what Burge (1996) himself has argued. In cases of self-knowledge, that subjects are the only ones to whom their contentful states appear gives them a *single*, unified perspective on those contents. Given that the content of the first-order state is the *same* content as that reviewed by the subject of the second-order state, and given that the perspectives of the subject as subject of the first-order state and as subject of the second-order state are *unified*, they cannot in general 'come apart': the thought by which a subject grasps a first-order state when thinking about it refers at the same time to it. This is not to say, however, that it is the *determinant* of it. It is to say that one thinks the thoughts one thinks one thinks.²²

²² Versions of this essay were read at Rutgers University and Temple University, and at a conference on Self-Knowledge in St Andrews, Scotland. I have benefited greatly from comments by those who attended these readings, especially Tyler Burge, Brian McLaughlin, Colin McGinn, Jerry Fodor, Brian Loar, and Tom Baldwin. I am also indebted to Graham Macdonald, Eve Garrard.

References

- Alston, W. (1971), 'Varieties of Privileged Access', *American Philosophical Quarterly*, 8: 223-41.
- Boghossian, P. (1989), 'Content and Self-Knowledge', *Philosophical Topics*, 17: 5-26.
- Burge, T. (1979), 'Individualism and the Mental', in P. French, T. Uehling, and H. Wettstein (eds.), *Midwest Studies in Philosophy* (Minneapolis: University of Minnesota Press), iv. 73-122.
- (1985a), 'Authoritative Self-Knowledge and Perceptual Individualism', in R. Grimm and D. Merrill (eds.), *Contents of Thought* (Tucson, Ariz.: University of Arizona Press), 86-98.
- (1985b), '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.

end p.153

- Burge, T. (1996), 'Our Entitlement to Self-Knowledge', *Proceedings of the Aristotelian Society*, 96: 91-116.
- Davidson, D. (1984), 'First Person Authority', *Dialectica*, 38: 101-11.
- (1987), 'Knowing One's Own Mind', *Proceedings and Addresses of the American Philosophical Association*, 60: 441-58.
- (1988), 'Reply to Burge', *Journal of Philosophy*, 85: 664-5.
- Descartes, R. (1969), *Meditations*, in M. Wilson (ed.), *The Essential Descartes* (New York, New American Library, Inc.).
- Heil, J. (1988), 'Privileged Access', *Mind*, 97: 238-51.
- (1992), *The Nature of True Minds* (Cambridge: Cambridge University Press).
- Macdonald, C. (1995), 'Externalism and First-Person Authority', *Synthese*, 104: 99-122.
- McDowell, J. (1994), *Mind and World* (Cambridge, Mass.: Harvard University Press).
- Moran, R. (1994), 'Interpretation Theory and the First Person', *Philosophical Quarterly*, 44: 154-73.
- Peacocke, C. (1992), *A Study of Concepts* (Cambridge, Mass.: MIT Press).
- (1996), 'Entitlement, Self-Knowledge and Conceptual Redeployment', *Proceedings of the Aristotelian Society*, 96: 117-58.
- Shoemaker, S. (1994), 'Self-Knowledge and Inner Sense', *Philosophy and Phenomenological Research*, 54: 249-314.
- Wright, C. (1987), 'Further Reflections on the Sorities Paradox', *Philosophical Topics*, 15: 277-90.

—— (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).