

INTROSPECTION, PERCEPTION, AND EPISTEMIC PRIVILEGE

Quassim Cassam

1

Could there be a creature whose knowledge of its own mental states or properties is perceptual and whose knowledge of the physical properties of external objects is introspective? The answer to this question obviously depends not only on how one conceives of the distinction between mental and physical properties but also on one's conception of the differences between perceptual and introspective knowledge. On one view, introspective knowledge enjoys a range of epistemic privileges which perceptual knowledge lacks. On this account, a creature whose knowledge of the physical properties of external objects is introspective would be one whose knowledge of such properties is epistemically privileged in the way that our knowledge of our own mental properties is epistemically privileged. By the same token, a creature whose knowledge of its own mental properties is perceptual would be one whose knowledge of its mental properties is subject to what McGinn describes as 'the frailties and fallibility of perceptual knowledge'.¹

What are the epistemic privileges of introspective knowledge, and what do the frailties of perceptual knowledge consist in? As McGinn admits, there are disagreements over the precise nature and extent of first-person authority, but his discussion assumes the strongest possible contrast between the two types of knowledge. His idea is that the special authority of introspective knowledge consists in its being 'non-criterial, non-inferential, direct, infallible, incorrigible, and certain'.² In addition, it is knowledge of facts that are 'self-intimating'. In contrast, perceptual knowledge is not authoritative. It is not knowledge of self-intimating facts, and the

knowledge that it furnishes us with is criterial, inferential, indirect, fallible, corrigible and uncertain.

McGinn's proposal is that 'the customary privileges of introspection can be inverted'.³ A fully inverted subject would be one whose knowledge of physical facts enjoys all the epistemic privileges of introspective knowledge and whose knowledge of mental facts enjoys none of these privileges. In arguing for this possibility, McGinn sees himself as undermining a 'venerable assumption'⁴ about the connections between introspection and the mental and perception and the physical. The venerable assumption is that these are necessary connections. Those who accept this assumption might be called traditionalists. As far traditionalists are concerned, it is not just that we actually introspect mental facts and perceive physical facts. It is also the case that these modes of access to the mental and the physical cannot be reversed. For McGinn, in contrast, there is no necessity here. We introspect mental facts and perceive physical facts, but there could be a creature that perceives mental facts and introspects physical facts. Matter cannot therefore be characterized as what is known about by perception and mind as what is known about by introspection.

How venerable is the assumption to which McGinn takes exception? It is worth remembering that introspection or consciousness has often been thought of as a form of inner perception. For a philosopher who agrees with Locke that conscious perception is a matter of perceiving that one perceives, the idea that mental facts can be perceived will not seem particularly radical or disconcerting.⁵ Such a philosopher might point out that there is a sense in which we actually perceive some of our own mental properties, so there is the simplest possible case for thinking that there could be a creature that perceives mental facts. We are such creatures. As for whether introspective knowledge of physical facts is possible, consider the case in which one

knows that one's legs are crossed as a result of feeling that one's legs are crossed. It is presumably a 'physical fact' that one's legs are crossed, but it is not implausible that one's knowledge 'from the inside' of the position of one's limbs is at least akin to introspective knowledge.⁶ So here is a case in which we actually have something like introspective knowledge of a physical fact.

Although it is worth drawing attention to these points, it is important that their significance is not exaggerated. To begin with, the idea that introspection is a form of inner perception is hard to defend. For if it were literally a form of inner perception, then it would be subject to all the frailties and fallibility of perceptual knowledge. So the fact that we do not regard it as subject to these frailties and fallibilities is itself a reason for denying that it is a form of inner perception. In contrast, one's knowledge from the inside of one's own body is subject to many, if not all, of the frailties and fallibility of perceptual knowledge, and this shows that our bodily self-knowledge is not literally introspective. After all, it is not as if judgements about the position of one's limbs are literally infallible, incorrigible or certain. Moreover, even if one were persuaded that our bodily self-knowledge is introspective, this would still not make it plausible that we have authoritative, introspective knowledge of the physical properties of genuinely external objects, that is, knowledge of the physical properties of objects which are distinct from our own bodies and which we do not feel in the way in which we feel our own bodies.⁷

These responses to the proposal that we actually perceive mental facts and introspect some physical facts rely on the assumption that there is the sharpest possible distinction between genuinely introspective and genuinely perceptual knowledge. One might wonder, therefore, whether this assumption is correct. Consider, once again, McGinn's list of the customary epistemic privileges of

introspective knowledge and the frailties of perceptual knowledge. Is it really plausible that genuinely introspective knowledge must enjoy all the privileges on this list and that genuinely perceptual knowledge can enjoy none of them? It remains to be seen what these privileges amount to, but it is not obvious that genuinely introspective knowledge cannot tolerate some degree of fallibility, or that genuinely perceptual knowledge must be inferential. Perhaps, in that case, more can be done with the idea that we actually perceive some of our own mental properties and introspect some of our own physical properties.

To argue in this way is not to collapse the distinction between introspective and perceptual knowledge, or to suggest that there is no sense in which the former is more authoritative than the latter. The issue is whether the contrast is as sharp or stark as has so far been assumed. Suppose, then, that one adopts a somewhat weaker conception of the contrast between the two types of knowledge. What would be the impact of this on McGinn's discussion? One thought that one might have is that the less demanding one's conception of the epistemic privileges of introspective knowledge the easier it should be to show that these privileges can be inverted. It is also arguable, however, that the weaker one's conception of the epistemic privileges of introspective knowledge the less startled one should be if it turns out that these privileges can be inverted, or that some of our existing knowledge of mental and physical facts is 'inverted'. It is understandable, therefore, that McGinn should choose to operate with an extremely demanding conception of the authority of first-person introspective knowledge. As he says, this conception makes the question of inversion difficult for him, but it also gives this question more bite than a weaker conception of the authority of introspective knowledge.

In the light of this discussion, a natural way of proceeding would be to start with a strong conception of the contrast between perceptual and introspective knowledge and to consider whether, on this conception, there could be a creature whose access to mental and physical facts is inverted in the way that McGinn proposes. I believe that when we understand what the epistemic privileges that McGinn lists really amount to, we will be in a position to see that they cannot be inverted. We can then ask whether the epistemic privileges of introspective and perceptual knowledge can be inverted on a weaker conception of the asymmetry between these two types of knowledge. I will argue, contrary to what might initially be supposed, that these privileges still cannot be inverted. In effect, therefore, I will be defending the venerable assumption to which McGinn takes exception. It will emerge as I go along that his major criticisms of this assumption depend on a reliabilist account of knowledge and justification, but this account fails to make sense of a good number of the epistemic privileges that have traditionally been associated with introspective knowledge. By the same token, it fails to make sense of the possibility that these privileges might be inverted.

2

Let us begin our examination of the privileges that McGinn attributes to first-person introspective knowledge by taking a closer look at the notions of infallibility, incorrigibility and certainty. Suppose that knowing that *p* involves believing that *p*. If *S*'s knowledge that *p* is to count as infallible, then it must be the case that *S*'s belief that *p* cannot be mistaken.⁸ For *S*'s knowledge that *p* to count as incorrigible, it must be the case that *S*'s belief that *p* cannot be corrected. To correct a belief is to show that it is mistaken. Thus, the incorrigibility of a belief consists in its being such that it cannot be shown to be mistaken. Finally, the certainty of a belief or piece of

knowledge may be supposed to consist in its being indubitable. One's belief that p is indubitable just if one could not have grounds for doubting it or for thinking that p is mistaken.

What, then, of the idea that one's knowledge of certain facts might be non-criterial, non-inferential and direct? Assuming that S cannot know that p unless S is justified in believing that p, S's knowledge that p will count as non-inferential just if S's belief that p is not justified on the basis of any other knowledge that S possesses. Direct knowledge, as McGinn understands it, is knowledge that involves no experiential intermediary. Equivalently, it is knowledge that is not mediated by sensory states. Finally, criterial knowledge is knowledge that is based on criteria. Roughly, a criterion for p is defeasible evidence for the truth of p. Thus, one's knowledge that p will count as non-criterial if and only if one's belief that p is not justified by defeasible evidence for the truth of p. In such cases, it must either be the case that one's belief is not justified by evidence, or that the justifying evidence is such that it cannot be defeated.⁹

The notion of self-intimation is somewhat different from the other notions considered so far. In the most basic sense of 'self-intimating', it is truths or facts which are self-intimating, rather than one's knowledge of truths or facts. For a fact to be self-intimating is for it to be such that it cannot obtain without one's knowing that it obtains. The implication is that self-intimating facts are ones with respect to which one cannot be ignorant. In Chisholm's words, such facts 'present themselves to the subject who has them'.¹⁰ They might therefore be described as self-presenting or as evident. Facts that are evident to one subject might, of course, not be evident to another. But if it is evident to S that p, then p cannot be true without S knowing that it

is true. To put it another way, it follows from the fact that p that S is justified in believing that p.

With these definitions in place, it is now possible to understand what traditionalists are committed to. One thing to which traditionalists are not committed is that one is authoritative in the strong sense with respect to all of one's own mental states or properties. Clearly, some of our mental properties are neither self-intimating nor such that our beliefs about them are infallible or incorrigible. To take a mundane example, one might be jealous of someone without realizing it, and one might falsely believe that one is not jealous or resentful of someone else's success.¹¹ On the other hand, there are other mental properties our knowledge of which traditionalists will want to regard as authoritative in the strongest possible sense. These are mental properties, perhaps like pains, which one cannot have without realizing that one has them. When one is in pain, it is evident to one that one is, and the introspectively based belief that one is in pain cannot be mistaken or corrected or open to doubt. Furthermore, one does not infer that one is in pain from something else that one knows, and one does not justify the belief that one is in pain by on the basis of defeasible evidence. Finally, one's knowledge or awareness that one is in pain involves no experiential intermediary. As Shoemaker points out, an object's appearing a certain way to someone involves that person's being in a subjective state with a certain phenomenal character but 'there is no such thing as a sense impression of a sensory state, having a phenomenal character of its own'.¹²

Needless to say, it is not uncontroversial that pains or any other mental properties actually have all of the mental properties that traditionalists attribute to them, but let us prescind for the moment, from further consideration of this issue. For McGinn is not primarily concerned to challenge the claim that our introspective

knowledge of our own mental properties is authoritative in the strong sense. His proposal is that there could be a creature whose knowledge of the external world of material objects is no less authoritative. So the question is whether this proposal is correct, that is, whether there could be a creature for whom the physical properties of external objects are self-intimating, and whose knowledge of such properties is non-criterial, non-inferential, direct, infallible, incorrigible, and certain.

In an attempt to make it plausible that there could be such a creature, McGinn introduces the notion of a 'P-state'. P-states are non-mental brain states that we come to know by means of perception, but McGinn sees no reason why it should not be possible for them to be known introspectively, with all the authority that springs from this mode of knowing. In McGinn's own words:

there is no deep conceptual obstacle to simply rigging up the brain in such a way that the presence of a P-state simply registers with the subject: he spontaneously finds himself with believing that the neurons in such-and-such area are firing at so-and-so rate.... The neural state in question has caused him to have a belief in its presence – directly, non-criterially, etc... The causal route simply doesn't go through a perceptual channel; there is no perceptual experience of the neural state – just an immediate conviction that it is there.¹³

Essentially the same story can be told about a creature's access to physical states outside the head:

Couldn't we in principle rig the world up in such a way that facts about physical objects are immediately fed into a person's brain and trigger beliefs in their existence, without any perceptual mediation...? Then we could describe this as a case in which either subjects are "introspecting" external objects or have a mode of access to them that mirrors introspection in central

respects.... The causal connection is totally reliable and the perceptual state has been eliminated from the process of belief formation: isn't this the essence of our knowledge of our own mental states?¹⁴

To sum up, the claim in these passages is that there could be direct, perceptually unmediated access to physical facts, and that to admit that this is possible is to admit that there could be introspective or quasi-introspective knowledge of physical facts.

Among several striking features of these passages, perhaps the most striking is the fact that they only pay lip service to several of the key epistemic privileges which McGinn attributes to ordinary introspective knowledge. The focus in these passages is on the directness of introspective knowledge, but what of the remaining privileges? If introspective knowledge must be infallible, then McGinn's story only makes it plausible that there could be introspective knowledge of physical facts if it makes it plausible that knowledge of physical facts could be infallible. The analogue for infallibility in McGinn's story is reliability, that is, the reliability of the process of belief formation, but a reliable process need not be one that is incapable of going wrong or of producing mistaken beliefs. Infallibility has to do with the impossibility of mistakes but the possibility of mistakes cannot be ruled out by stipulating that the inverted subject's beliefs about external objects are reliable.

It might be objected that this difficulty can be sidestepped by stipulating that a subject with introspective access to physical facts is one whose beliefs about external objects are not just reliable but totally reliable. Totally reliable beliefs about external objects 'just pop into your head and they are (almost?) never wrong'.¹⁵ But this still falls short of infallibility as the traditionalist understands it. A totally reliable process or mechanism is one which never goes wrong in relevant contexts, not one which could not possibly go wrong, or which could not cease to be reliable.¹⁶ McGinn's

discussion assumes ‘the correctness of some sort of reliability account of knowledge and justification’,¹⁷ and one of the central aims of such accounts is to make knowledge naturalistically respectable. It seems unlikely, however, that any naturalistically respectable notion of total reliability can provide for the kind of infallibility that the traditionalist is after.

Moreover, even if one is prepared to grant that total reliability amounts to infallibility, it is a further question whether it amounts to incorrigibility and certainty. Taking these in reverse order, one’s beliefs about external objects are certain only if there can be no grounds for doubting them, but a subject whose beliefs about external objects are always correct might still have grounds for doubting whether they are always correct. Equally, a subject whose beliefs about external objects are reliable might, in certain circumstances, be forced to correct an individual belief of this type. Unless the mechanisms of belief formation are such that they could not possibly malfunction, it is difficult to see how the inverted subject’s beliefs about external objects could be literally incorrigible.

In addition to these limitations, there is also an obvious difficulty with the idea that physical facts are evident to the inverted subject. Self-intimating or evident facts are ones that cannot obtain without the subject knowing that they obtain, but there will presumably be many physical facts that are not, in this sense, evident to the inverted subject. Even in the case of those physical facts with respect to which the inverted subject is not ignorant, it would be difficult to maintain that these facts are evident to the subject. Evident or self-intimating facts are ones that present themselves to the subject, and this is a matter of the subject’s being aware that these facts obtain. In contrast, to say that beliefs about external objects just pop into the subject’s head is not to say that the corresponding facts present themselves to the

subject. As McGinn remarks, the inverted subject is like someone with blindsight, but the important point about blindsight is that those suffering from this condition are able to form true beliefs about facts that are, in an important sense, not evident to them.

The remaining epistemic privileges of introspection are less difficult for McGinn to accommodate. If the inverted subject's beliefs about external objects amount to knowledge, then the fact that the knowledge in question is not perceptually mediated shows that it is direct. It is also true that the inverted subject's beliefs about external objects are not justified by any of its other beliefs, and that they are not justified by defeasible evidence. So the inverted subject's knowledge is non-inferential and non-criterial, as well as direct. This might make it seem that proponents of the view that there could be introspective knowledge of physical facts can claim at least a partial victory, but such an impression would be misleading. In the first place, there is a question about whether the inverted subject's beliefs about external objects really do amount to knowledge. By McGinn's lights, these beliefs are justified and amount to knowledge because they are reliable. This assumes the correctness of a reliabilist account of knowledge and justification, and so ought not to satisfy traditionalists who are not reliabilists. Indeed, one might think that the lesson of the discussion so far is that reliabilism cannot make sense of a good number of the epistemic privileges of genuinely introspective knowledge, and that this is what ultimately accounts for failure of McGinn's attempt to represent the inverted subject's knowledge of physical facts as genuinely introspective.

Secondly, and more importantly, the fact that the inverted subject's knowledge of physical facts is direct, non-inferential and non-criterial only supports that proposal that its knowledge is introspective rather than perceptual if we take it that perceptual knowledge is not direct, non-inferential and non-criterial. Although this is precisely

what McGinn's discussion assumes, there is actually very little to be said for this assumption. Criterial knowledge is knowledge that is based on defeasible evidence, and if the belief that *p* is supported by defeasible evidence then it is possible that not-*p*.¹⁸ But if one genuinely sees that *p* (where *p* is a proposition about an external object), then it not possible that not-*p*.¹⁹ Seeing that *p* is the case is, in this sense, a way of coming to know that *p* is the case, but this knowledge is both perceptual and non-criterial. It is also non-inferential, to the extent that it is one's perceptual awareness of *p* that justifies the corresponding belief. In this case, one's belief that *p* is not justified by anything else which one knows or believes.

There remains the question of directness. If, as McGinn maintains, the positing of experiential intermediaries in perceptual knowledge carries no commitment to a sense-datum theory of perception and is compatible with naïve realism, then one can perhaps agree that perceptual knowledge is indirect. All this means is that one cannot have perceptual knowledge without having perceptual experiences, and this seems an innocuous thought as long as care is taken not to build too much into the notion of a perceptual experience. Since the inverted subject's knowledge of physical facts is not experiential knowledge, it is not perceptual knowledge and has at least this much in common with genuinely introspective knowledge. This is not to say, however, the inverted subject's knowledge of physical facts is genuinely introspective. It cannot be genuinely introspective without the other epistemic privileges of such knowledge.

To sum up, what McGinn needs to establish is that his inverted subject would be authoritative in a very strong sense about the physical properties of external objects, but what he actually establishes is something much weaker than this. The most that he establishes is that an inverted subject's 'knowledge' of physical facts would be reliable and direct, but reliability and directness do not, on their own,

constitute ‘the essence of our knowledge of our own mental states’.²⁰ As far as traditionalists are concerned, the essence of our knowledge of our own mental states is its infallibility, incorrigibility and certainty, but McGinn makes no serious attempt to demonstrate that an inverted subject would enjoy these epistemic privileges with respect to the physical properties of external objects. He therefore fails to make a convincing case for the proposal that external objects can be known with all the authority accorded to introspection. Furthermore, the epistemic privileges which an inverted subject would enjoy with respect to such properties include ones which also characterize ordinary perceptual knowledge of external objects, and this strengthens the case for resisting the idea that an inverted subject’s epistemic access to physical facts would be introspective. The discussion so far suggests that the access that McGinn describes would be neither introspective nor perceptual. Indeed, from a non-reliabilist perspective, it is not beyond dispute that an inverted subject would have knowledge of physical facts.

How, then, do traditionalists themselves account for what they regard as the authority of first-person introspective knowledge? Many different explanations have been proposed, but one historically influential line of thought claims that it is in the nature of some mental properties that they are transparent to their subject, and that it is in virtue of their transparency that one’s introspectively based judgements about such mental properties cannot be mistaken, corrected or doubted. This appears to be Hume’s proposal in the following well-known passage:

For since all actions and sensations of the mind are known to us by consciousness, they must necessarily appear in every particular what they are, and be what they appear. Everything that enters the mind, being in reality a perception, ‘tis impossible anything should to feeling appear different. This

were to suppose that even when are most intimately conscious, we might be mistaken.²¹

Here, it is the transparency of perceptions, the fact that they must necessarily appear what they are and be what they appear, which is held to account for the impossibility of mistakes when we are ‘most intimately conscious’. Although this does not address the incorrigibility and certainty of first-person introspective knowledge, Hume would doubtless regard these epistemic privileges as flowing from the transparency of the mental. As for the idea that mental properties are self-intimating, this can be seen as a direct consequence of their transparency.

On a Humean account of first-person authority, it is not difficult to see why the idea that such authority might attach to a creature’s knowledge of physical facts would seem so problematic. For the physical properties of external objects are, by their nature, not such that they must necessarily appear what they are and be what they appear. That is why judgements about external objects can be mistaken, and why scepticism is a genuine issue in this context. Unless, therefore, it makes sense to suppose that physical facts could be as transparent to the inverted subject as mental properties are to us, it does not make sense to suppose that an inverted subject’s knowledge of physical facts could be as authoritative as our knowledge of certain mental facts. The traditionalist’s claim is that it does not make sense to suppose that physical facts could be as transparent to the inverted subject as mental properties are to us, and that the reliability of the inverted subject’s judgements about the physical properties of external objects is no substitute for this transparency.

From McGinn’s perspective, these remarks might seem question begging. It might be held that the claim that mental properties must be transparent and that physical properties cannot be transparent in the same way is just the claim that

McGinn is attempting to question, and that this epistemological asymmetry between mental and physical facts cannot therefore be assumed as a premise in an argument against the possibility of inversion. In fact, this is not the point. The point is that if we want to understand how external objects can be known with all the authority accorded to introspection, we must first understand the nature and grounds of this authority. On a traditionalist view, the nature and grounds of first-person authority are such that one's knowledge of the physical properties of external objects could not be authoritative.²² It is open to McGinn to reject this account, but this will only advance matters if he has a viable alternative. With an alternative account of the authority of first-person introspective knowledge in place, we can test whether this authority can be inverted. The problem, I have suggested, is that McGinn lacks a viable account of the authority of first-person introspective knowledge, and this makes it seem that the inversion hypothesis is much easier to sustain than it really is. If we understand first-person authority in reliabilist terms, we should not be surprised that the authority of first-person introspective self-knowledge can be inverted. But if we understand first-person authority in reliabilist terms, then it will also be the case that we will not be able to do justice to the nature and extent of this authority.

3

So far, I have only commented on one half of the inversion hypothesis. What about the other half, the idea that there could be a creature whose knowledge of its own mental states would be subject to all the frailties and fallibility of perceptual knowledge? We ascribe pain to others on the basis of behavioural evidence, but the inverted subject 'never ascribes the concept pain on any other basis than behaviour'.²³ Even when it comes to ascribing pain to itself, it relies on behavioural evidence. This means that its self-ascriptions of pain will be fallible, corrigible and uncertain, in the

way that our ascriptions of pain to others on the basis of behavioural evidence are fallible, corrigible and uncertain. Since the behavioural evidence that the inverted subject relies on is defeasible, its knowledge of its own mental properties will be a form of criterial knowledge, and there is also a case for regarding it as inferential and indirect. In these respects, its self-knowledge will be subject to what McGinn regards as all the usual errors of perceptual knowledge.

For the purposes of this argument, it is important that the inverted subject is thought of as operating with the same mental concepts that we operate with, and this seems doubtful on an epistemic view of concepts. But McGinn rejects this view of concepts. He thinks that concepts are individuated not by their conceptual role or the epistemic faculties by means of which they are applied but by their reference. Thus, the use of a different evidential basis for applying a concept does not create a new concept, and ‘the inverted subject has the same mental concepts as us, since his concepts refer to what our concepts refer to’.²⁴ Indeed, the fact that the inverted subject can apply the concept pain to itself on the basis of behavioural evidence is seen by McGinn as bringing out what is wrong with views that tie concepts to epistemic capacities.

This argument raises many questions, but perhaps the most obvious concerns the role of phenomenology in the inverted subject’s self-ascriptions of pain. McGinn’s idea is that pains do not lose their phenomenal character under inversion, since their phenomenology is not a function of the way in which one has epistemic access to them. This makes it possible for him to assert that in the inversion scenario ‘the sensation of pain is still phenomenally individuated, even though it is not ascribed on the basis of its phenomenology’.²⁵ In response to the worry that the phenomenology of pain cannot be separated in this way from the basis upon which pains are ascribed,

McGinn points out that this is exactly what happens with third-person ascriptions. When one ascribes pain to someone else, one does not do so on the basis of phenomenology, but what one ascribes is nevertheless phenomenally individuated.

Among the paradoxical consequences of this view, one is that it opens the door to a form of scepticism with respect to knowledge of one's own mind that is analogous to scepticism about other minds. Although McGinn is happy to concede this point, it is harder to swallow than his discussion suggests. When one thinks about other minds, one familiar thought is that someone else might be in pain without manifesting it behaviourally or, more controversially, without even being disposed to manifest it behaviourally.²⁶ Another familiar thought is that someone who appears to be in pain as far as her behaviour goes might not really be in pain. The conclusion that sceptics about other minds draw from these familiar thoughts is that other minds are, in these respects, opaque to one. So if the inversion scenario makes room for scepticism with respect to knowledge of one's own mind, then it ought to be possible to think of one's own mind as, for similar reasons, no less opaque to one. The question that needs to be addressed, therefore, is whether this is indeed a genuine possibility, given that that we are presently concerned with mental properties that are phenomenally individuated.

Consider the first kind of opacity. Suppose that the inverted subject is in a state that has the phenomenal character of pain but stoically displays no pain behaviour and is not even disposed to display such behaviour. In a case like this, the subject is still entitled to ascribe pain to itself just on the basis of its phenomenal character, just on the basis of how it feels; the absence of behavioural evidence will not stymie the inverted subject's self-ascriptions of pain in the way that it might stymie its other-ascriptions of pain. Similarly, as far as the second kind of opacity is

concerned, the inverted subject might wonder whether another subject who displays pain behaviour is really in pain, but it cannot intelligibly wonder whether it is really in pain when it displays such behaviour. If, as McGinn concedes, ‘the feeling of pain cannot be pulled apart from pain itself’,²⁷ and the inverted subject feels pain, then it is in pain and knows that it is (assuming that it has the necessary conceptual capacities). It cannot deceive itself with respect to its own phenomenally individuated states in the way in which it can deceive others with respect to such states.

The point of these considerations is not to suggest that we could not, in peculiar or unusual circumstances, self-ascribe pain on the basis of behaviour. The point is rather that pains can always be adequately self-ascribed on the basis of how they feel, and that this is the canonical basis for self-ascribing pains. Against this background, it is difficult to make sense of the idea of an inverted subject who never self-ascribes pains on the basis of anything other than behaviour. It is worth adding that this idea is made no less paradoxical by the observation that we do not actually ascribe pains to others on the basis of phenomenology. For what we ascribe when we ascribe pains to others are precisely states that we suppose they can ascribe to themselves on the basis of phenomenology. We understand how we might mistakenly believe that others are in pain or fail to realize that they are in pain, but we do not understand how they might mistakenly believe that they are in pain or fail to realize that they are in pain. What this suggests is that it is not an acceptable consequence of inversion that it generates, or purports to generate, a form of scepticism about knowledge of one’s own mind that parallels a familiar form of scepticism about other minds.

This argument might be accused of begging the question against McGinn’s proposal. What I have been assuming is, in effect, that pains and some other mental

properties are necessarily self-intimating, yet one of the aims of McGinn's discussion is to dispute this traditional assumption. If pains are necessarily self-intimating, then it follows that they cannot be opaque to their subject or canonically self-ascribed on the basis of behavioural evidence. But if the inverted subject self-ascribes pains on the basis of behavioural evidence, and can wonder whether it is really in pain, then what this goes to show is that pains are not necessarily self-intimating. To assume otherwise is, according to the present line of thought, to reject McGinn's position rather than to argue against it.

In a way, this is right, but it is difficult to assess the proposal that a creature could have inferential access to its own pains without making any assumptions about what pains fundamentally are. Without a conception of what pains fundamentally are, it becomes indeterminate whether what is being envisaged in the inversion scenario is a creature which is self-ascribing pains on the basis of behavioural evidence or self-ascribing some other kind of mental property. Traditionalists claim that what makes a pain a pain is not just its phenomenal character but the fact that it is evident to its subject and so can be known about in certain ways. Indeed, they see the fact that one has non-inferential access to one's own pains as an aspect, or at least as a consequence of, their phenomenal character. McGinn is committed to rejecting this account of the intrinsic nature of pain, but it is not clear what he puts in its place. He says that the phenomenal appearance of pain is intrinsic to it, but the question which this not address is whether the phenomenal appearance of pain can be understood in terms which to do imply that pains are evident their subject. For this reason, the insistence that that the properties which the inverted subject self-ascribes on behavioural grounds are really pains will carry little weight with McGinn's traditionalist opponents.

These concerns about McGinn's position are related to concerns that one might have on an epistemic view of concepts. Just as one might wonder whether it is really pains that the inverted subject ascribes to itself on the basis of behavioural evidence, so one might wonder whether the inverted subject is really operating with the concept pain in any of its self-ascriptions. Perhaps these are just two different ways of making the same point. As remarked above, McGinn's view is that the inverted subject has the same mental concepts as us because its concepts refer to what our concepts refer to, but an account still needs to be given of what makes it the case that the inverted concept of pain and our concept of pain refer to the same property. The traditionalist's proposal is that reference is determined by conceptual role, and that our concept of pain and the inverted subject's concept cannot therefore have the same reference. If something along these lines can be made out, then this strengthens the case for denying that the inverted subject's self-ascriptions are, or include, self-ascriptions of pain.

The lesson of the discussion so far is that the kind of inversion which McGinn envisages is much more difficult to think through than one might suppose. We saw above that there could not be a creature whose knowledge of external objects is as authoritative as our introspective self-knowledge. At any rate, McGinn has not shown that that could be such a creature. We have now seen that there could not be a creature whose knowledge of its own mental properties is as unauthoritative as our perceptual knowledge or our knowledge of other minds. Indeed, the inverted subject's self-knowledge would be, in some ways, less authoritative than ordinary perceptual knowledge, but self-ascriptions of pains and other phenomenally individuated mental properties cannot be less authoritative than ordinary perceptual knowledge. Since a fully inverted subject would be one with both authoritative knowledge of external

objects and unauthoritative knowledge of its own mental properties, these considerations add up to a strong case against the possibility of full inversion.

One might wonder, however, how much of what I have been arguing relies upon what some would regard as an excessively demanding conception of first-person authority. Would it be easier to make a case for the possibility of full inversion on a less demanding conception of first-person authority? I would like to bring this discussion to a close by examining the prospects for a watered down conception of first-person authority, and explaining why the argument for full inversion remains problematic on any plausible conception of such authority. The point at which full inversion becomes viable is the point at which the notion of first-person authority is watered down beyond all recognition. Unfortunately, this is also the point at which the possibility of full inversion becomes relatively uninteresting.

4

I have suggested that traditionalists regard infallibility, incorrigibility and certainty as the essence of our knowledge of our own mental states. Yet there are clearly some errors to which even our introspective knowledge is not immune. One can misdescribe mental properties to which one has privileged access, and there can also be errors that are caused by carelessness or inattention. This suggests that a realistic conception of first-person authority will not insist that authoritative knowledge must be completely immune to error. What is special about authoritative knowledge is that it is immune to certain kinds of error. As long as our introspective knowledge is not subject to errors of these specific kinds, it still counts as authoritative.

Although the concession that authoritative knowledge need not be completely immune to error might be thought to represent a weakening of the notion of first-

person authority, it is worth remembering that many traditionalists have been prepared to make this concession. Their idea is that the sense in which introspective knowledge is infallible is that one's introspective judgements cannot be mistaken as long as one attends to their subject matter with sufficient care and attention. In Burge's terminology, mistakes which are the result of carelessness or inattention 'indicate something wrong with the thinker',²⁸ and so are not 'brute errors'.²⁹ In these terms, part of what makes introspective knowledge authoritative is that it does not allow for the possibility of brute error. In contrast, perceptual knowledge is not authoritative because a person 'can be perceptually wrong without there being anything wrong with him'.³⁰

On this account of the contrast between introspective and perceptual knowledge, can their epistemic privileges be inverted? Let us go back to the idea that an inverted subject is one whose beliefs about external objects are reliable. It was suggested above that even a creature that is as reliable about the properties of external objects as McGinn stipulates can cease to be reliable. One way in which this might happen is if some of its cognitive mechanisms develop a fault and start to malfunction. In the event of a malfunction, it would cease to be reliable in McGinn's sense and its beliefs would no longer be immune to error. Would such errors be brute? Assuming that they are not the result of carelessness or inattention, one might think that they would be brute errors. But a malfunction in its cognitive mechanisms surely indicates that there is something wrong with the subject, and this is at odds with the idea that its errors would be brute.³¹ So even if one agrees with Burge that 'in all cases of authoritative knowledge brute mistakes are impossible',³² the fact that the mechanisms that lead to the formation of beliefs about external objects can

malfunction does not show that the inverted subject's knowledge of external objects would not be authoritative.

In response to this line of argument, traditionalists should question the lumping together of carelessness, inattention and malfunction as different sources of error which are all compatible with authoritative knowledge. It is true that these sources of error all indicate that there is something wrong with the subject, but they do so in significantly different ways. For a thinker can be blamed for carelessness or inattention but not for having malfunctioning cognitive mechanisms. In addition, errors induced by cognitive malfunctions are, from the subject's own perspective, no less brute than other paradigmatically brute errors. These considerations imply that the inverted subject's knowledge of external objects is not genuinely authoritative, in so far as its beliefs about external objects can be brutally mistaken. In the case of genuinely authoritative knowledge, in contrast, not even a malfunction can cause one to go wrong.

The lesson of this discussion is that the inverted subject is going to count as authoritative about external objects only if the notion of authoritative knowledge is watered down to the extent of allowing that such knowledge need not be immune to errors induced by cognitive malfunctions. Even then problems would remain for the hypothesis of full inversion. Physical facts would still not be evident to the inverted subject, and there would be important differences between the sense in which our beliefs about our own mental properties are reliable and the sense in which the inverted subject's beliefs about external objects would be reliable. In our case, it is not just that we never make substantive mistakes about our mental properties. Since these properties, or some of them at least, are self-intimating, it is evident to us how our beliefs about them are accurate, and it is also evident to us that they are accurate. In

contrast, merely having accurate beliefs about external objects popping into one's head provides one with no insight whatsoever into either the source or the extent of their accuracy. It is difficult enough to sustain the idea that such beliefs amount to knowledge, let alone that they amount to authoritative knowledge.³³

Where does this leave McGinn's proposal? We have seen that McGinn fails to make it plausible that a creature's knowledge of the physical properties of external objects could be authoritative in a strong sense of 'authoritative'. On a weaker conception of authoritative knowledge, say one that allows for errors caused by cognitive malfunctions, the prospects for his proposal are slightly better but still not good. On any conception of first-person authority which even comes close to doing justice to the all the different senses in which ordinary introspective knowledge has traditionally been regarded as authoritative, there is no such thing as authoritative knowledge of external objects. There might be elements of traditional conception of authoritative knowledge which are negotiable, and one might be able to circumvent a good many of the difficulties to which I have been drawing attention by substantially modifying the traditional conception. It might be suggested, for example, that it is not just the idea that authoritative knowledge is completely infallible which needs to be questioned but also the idea that it is knowledge of self-intimating facts. But if one were that sceptical about the traditional conception of authoritative knowledge, then it is not clear why one would still be interested in discovering whether the epistemic privileges of introspection and perception can be inverted. In sum, the assumptions about the nature of authoritative knowledge that make McGinn's proposal interesting also make it unworkable.³⁴

¹ Colin McGinn, “Inverted First-Person Authority” (typescript), p. 2.

² “Inverted First-Person Authority”, p. 4.

³ “Inverted First-Person Authority”, p. 2.

⁴ “Inverted First-Person Authority”, p. 2.

⁵ Locke talks about perceiving that one perceives in Book II, chapter xxvii, section 9 of his Essay Concerning Human Understanding (Oxford: Oxford University Press, 1975). Perceiving the operations of our own minds is ‘reflection’, and involves the exercise of what Locke calls ‘internal Sense’ (Essay, II. i. 4).

⁶ For a defence of this idea, see my “Introspection and Bodily Self-Ascription”, in José Bermúdez, Anthony Marcel and Naomi Eilan (eds.) The Body and the Self (Cambridge, MASS: MIT Press, 1995), pp. 311-36.

⁷ This is not to deny that there is a sense in which one’s own body is an external object.

⁸ There is a useful account of the notion of infallibility and related terms of epistemic appraisal in William Alston, “Varieties of Privileged Access”, American Philosophical Quarterly, 8 (1971), pp. 223-41. My discussion is indebted to this paper, although some of the definitions that follow are different from Alston’s.

⁹ One might think that it is in the nature of evidence that it can be defeated, although we do speak of ‘incontrovertible’ evidence.

¹⁰ Roderick M. Chisholm, The Foundations of Knowing (Minneapolis: University of Minnesota Press, 1982), pp. 9-10.

¹¹ Some traditionalists would argue that one cannot fail to realize that one is jealous if one is sufficiently careful and attentive. For more on this idea, see section 4 below.

¹² Sydney Shoemaker, ‘Introspection and the Self’, in Quassim Cassam (ed.), Self-Knowledge (Oxford: Oxford University Press, 1994), 118-39, p. 135.

-
- ¹³ “Inverted First-Person Authority”, p. 3.
- ¹⁴ “Inverted First-Person Authority”, p. 4.
- ¹⁵ “Inverted First-Person Authority”, p. 4.
- ¹⁶ In a certain sense, therefore, there is no such thing as total reliability.
- ¹⁷ “Inverted First-Person Authority”, n. 15.
- ¹⁸ Cf. McDowell, ‘Criteria Defeasibility, and Knowledge’, in John McDowell, Meaning, Knowledge and Reality (Cambridge, MASS: Harvard University Press, 1998), pp. 369-94.
- ¹⁹ In Williamson’s terminology, ‘seeing that p’ is ‘factive’. See Timothy Williamson, “Is Knowing a State of Mind?”, Mind, 104 (1995), 533-65.
- ²⁰ “Inverted First-Person Authority”, p. 4.
- ²¹ David Hume, A Treatise of Human Nature, edited by L. A. Selby-Bigge (Oxford: Clarendon Press, 1978), p. 190.
- ²² As traditionalists see things, the fact that our knowledge of physical properties could not be authoritative is a reflection of the fact that these properties are objective, in a way that mental properties are not.
- ²³ “Inverted First-Person Authority”, p. 4.
- ²⁴ “Inverted First-Person Authority”, p. 5.
- ²⁵ “Inverted First-Person Authority”, p. 5.
- ²⁶ The coherence of this thought might be questioned by Wittgensteinians, but I do not propose to go into this here.
- ²⁷ “Inverted First-Person Authority”, p. 6.
- ²⁸ Tyler Burge, “Individualism and Self-Knowledge”, in Quassim Cassam (ed.) Self-Knowledge (Oxford: Oxford University Press, 1994), 65-79, p. 74.
- ²⁹ “Individualism and Self-Knowledge”, p. 73.

³⁰ “Individualism and Self-Knowledge”, p. 73. Burge regards misperceptions and hallucinations as brute perceptual errors.

³¹ In Burge’s view, ‘brute errors do not result from any sort of carelessness, malfunction, or irrationality on our part’ (“Individualism and Self-Knowledge”, p. 73).

³² “Individualism and Self-Knowledge”, p. 74.

³³ They would certainly not amount to what Michael Ayers calls ‘primary knowledge’. When we know something in this sense, ‘we know how we know it’ (Michael Ayers, Locke, volume I: Epistemology (London: Routledge, 1991), p. 140).

³⁴ I thank Ciara Fairley for helpful comments and discussion.