

Chapter 5

Introduction

I have been arguing that phenomenology can help us to diagnose why naturalistic accounts of the conscious mind run into an explanatory gap. My aim is to argue for the possibility of a naturalised phenomenology which can help the naturalist to close this gap. At the end of the last chapter a worry was raised that the phenomenological account of why things feel and seem as they do leaves untouched a central question that naturalists have addressed when thinking about the explanatory gap. The ideas I have introduced from phenomenology do not explain why there should be something rather than nothing it is like for a subject to undergo a conscious experience. This chapter will be concerned with introducing the phenomenologist's answer to this question.

I shall describe experiences that seem or feel a certain way to their subjects as experiences that have "phenomenal character". The question this chapter has as its concern asks what it is for an experience to have a phenomenal character. In section 1 I propose an answer to this question. There I will argue that an experience seems or feels a certain way to a subject when a subject can have knowledge of an experience he is undergoing from the inside without having recourse to observation or inference. I shall call knowledge of this kind "first-person knowledge". A subject can have first-person knowledge of his experiences I shall claim because they seem or feel a certain way to him. Thus an answer to the question of what it is for an experience to have a phenomenal character, I will suggest, lies with an explanation of how we can have first-person knowledge of our experiences.

In section 2 I set out two conceptions of phenomenal character both of which are agreed that phenomenal properties are nothing over and above representational properties. The first claims that phenomenal character is something *of* which a subject is conscious. The second denies that phenomenal character is something of which the subject is conscious, claiming instead that it is something *with* which the subject is

Consciousness, Self-Consciousness and First-Person Knowledge

conscious. We shall see that the first conception of phenomenal character marries up with higher-order theories of consciousness which claim that an experience is phenomenally conscious only if the subject represents that she is undergoing this experience.¹ The second conception of phenomenal character fits with an intentionalist account of phenomenal consciousness. I shall argue that standard intentionalist accounts of consciousness do not give us a satisfying answer to the question of how we can know our experiences from the inside. I look to an account of Brentano to see if he succeeds in finessing the problem. However I find his account wanting.

In section 3 I introduce an intentionalist account of phenomenal consciousness which originates in Husserl, and was developed in greater depth by Sartre and Merleau-Ponty. It claims that every conscious experience has built into it a variety of self-consciousness. Sartre calls the kind of self-consciousness in question “pre-reflective self-consciousness” to distinguish it from self-consciousness that is the outcome of introspective reflection. It is this second proposal which I will argue supplies the intentionalist with an account of phenomenal character. On this view an experience has phenomenal character when it is pre-reflectively self-conscious.

By section 4 we will have two distinct accounts of phenomenal character before us. The first proposal defended by higher-order theories of consciousness says an experience has phenomenal character when a subject represents herself having this experience. The second says an experience has phenomenal character when this experience has built into it pre-reflective self-consciousness. Section 4 presents an argument from Sartre for the conclusion that conscious experiences are also pre-reflectively self-conscious. If successful this argument will serve to further motivate the phenomenologist’s account of phenomenal character over the rival proposal of the higher-order theorist. We will see that Sartre’s argument succeeds against higher-order theories of consciousness only by

¹ Higher-order theories come in different forms. Lycan (1996) defends a version of higher-order perception theory. Rosenthal (1986; 1990/1997 & 1993) and Carruthers (2000) defend different versions of higher-order thought theory.

Consciousness, Self-Consciousness and First-Person Knowledge

assuming that there are no unconscious mental states. This is of course an assumption that higher-order theories reject.

In section 5 I offer an alternative argument against higher-order theories based on Sartre's argument. Sartre's argument is directed at theories which equate self-consciousness with self-knowledge. I argue that higher-order theories fit this description, and that there are certain features of self-consciousness which theories of this kind cannot accommodate. I conclude by showing how the account of phenomenal character I have proposed helps us to clarify the charge that naturalists cannot admit the existence of subjective facts.

1. Pinpointing the Problem

The question that will occupy us in this chapter is why it should be that there is something rather than nothing it is like for a subject to undergo a conscious experience. The difference between experiences there is something it is like to have, and experiences there is nothing it is like to have, turns on the presence or absence of a certain kind of consciousness I have been calling "phenomenal consciousness". I shall begin by considering how to draw the contrast between those states that are phenomenally conscious and those that are not.

This problem is particularly urgent for any intentionalist account of phenomenal properties which says that an experience seems or feels a certain way for a subject in virtue of its representational properties. All kinds of systems and devices represent the world as being a certain way that lack phenomenal consciousness. My homeostatic system represents the level of glucose in my blood, but the representations it produces will not be ones which make me conscious of anything. What distinguishes cases of representation like this from the kinds of representational states that do make me conscious of something? By endorsing intentionalism I have denied that there is anything over and above an experience's representational properties that gives an experience its

Consciousness, Self-Consciousness and First-Person Knowledge

phenomenal character. This leaves me needing to explain what differentiates those representational properties that give a representational state its phenomenal character from those that do not. The objection raised at the end of the last chapter was that so far, the phenomenologically-inspired version of intentionalism I have proposed hasn't given us an answer to this question.

I believe an answer to our question lies in the peculiar kind of first-person knowledge a subject can have of his phenomenally conscious mental states. A subject can know what it is like to undergo an experience from the inside, which is to say that a subject can know what an experience is like just by having that experience. If an experience doesn't have any phenomenal character, a subject cannot know the experience from the inside – he cannot know the experience just by having it. Perhaps we can say that the difference between representational states that make a creature conscious of something and those that do not lies in the epistemic access we have to the former but not the latter. We can say that there is something it is like for a creature to be in a representational state R when a creature is in a position to know that it is in R “from the inside”, just by tokening R. There is nothing it is like for a creature to be in a representational state R when a creature cannot have this kind of first-person knowledge of R.²

I shall say that a creature can have first person knowledge of a representational state R if R meets the following two, closely related, conditions:

- (1) A creature X can have *direct* knowledge that it is in R without first having to form observations either of itself, or of anything else, and

² Some might complain that this proposal ignores Block's (1995/1997) distinction between “access” consciousness and “phenomenal consciousness”. A state is access conscious for Block if ‘it is poised for free use in reasoning and for direct “rational control” of speech and action’. (*ibid*, 382) I have stipulated that something akin to this condition must be met if an experience is to have *phenomenal* character. The only way I can see to defend Block's distinction is to accept the existence of what Block calls “p-conscious properties” which are distinct from any cognitive, intentional or functional property (see Block, 1995/1997: 381). In chapter 4 I proposed a version of intentionalism which denies that p-conscious properties are distinct from intentional properties. Thus I reject Block's distinction because it seems to require a commitment to phenomenal properties conceived of as something over and above an experience's intentional properties.

Consciousness, Self-Consciousness and First-Person Knowledge

(2) A creature *X* can have *immediate* knowledge that it is in *R*, which is to say that *X* can know that it is in *R* without performing any conscious act of inference.

First-person knowledge is introspective knowledge: the kind of knowledge a subject can achieve by deliberately reflecting on or paying attention to a thought or experience he is having at the time.³ When we focus on an itch, tingle or pain we thereby make the sensation we are undergoing into the object of our reflection. Normally, we will in this way arrive at knowledge of what we feel: we will come to know that we are undergoing a particular sensory experience. In a similar fashion we can make a belief the object of an act of reflection: I might consider for instance whether I believe that Tony Blair can be trusted. By considering my attitude to this proposition I come to know something about myself, I come to know whether this is something I believe.

It is only my own experiences which I can know in the first-person: no other subject can know my experiences from the inside, and nor can I know of another subject's experiences in this way.⁴ The difference between first-person knowledge and what I shall call "third-person knowledge" can be traced to a difference in the warrant I have for these respective types of knowledge-claims. The warrant I have for my first-person knowledge-claims derives from my experience's phenomenal character, from the way my experiences present the world as seeming to me. What I know directly and immediately is the way an experience presents the world as seeming. The warrant I have for the claims I make about the experiences of another person, come from the observations I form of that

³ Whether this act of reflection is best understood as a kind of inner-perception is a question I shall ignore for now. For a defence of an understanding of introspection along these lines see Armstrong (1968) and for some trenchant and to my mind persuasive criticism see Shoemaker (1994/1996, lecture 2). I will postpone discussion of this question till later in section 5.

⁴ This is certainly not to deny that I can have knowledge of another subject's experiences. What I cannot have is first-person knowledge of their experiences. I will call the knowledge I can of another person's mind, "third-person knowledge". We shall see later that third-person knowledge is distinguished from first-person knowledge by the kind of warrant I have for assertions made in the first-person.

Consciousness, Self-Consciousness and First-Person Knowledge

person and the inferences I make based on those observations. I can know I am feeling silk just by having an experience of the feel of the silk, but I cannot know that another person is feeling silk except by seeing them pick it up and feel it or by having them tell me this is what they are experiencing. I cannot know the experiences of another person from the inside; only they can know their experiences in this way. Hence the claims we make about another's persons experiences do not derive their warrant from an experience's phenomenal character in the way that the claims we make about our own experiences do.

I have proposed that we use the first-person knowledge a subject can have of his experiences as an initial way of distinguishing those representational states that have a phenomenal character from those that do not. We can now see how this might work. The states which I can know directly and immediately are states which I can know from the inside just by having them. I can know of these states just by having them because these states have phenomenal character. It is because there is a way these states seem or feel to me that I can know them from the inside just by having them. If there is no way a state seems or feels to me when it occurs I will not be able to know it in this way, just by having it.

If we combine this suggestion with the intentionalist thesis that the way an object seems to a subject is fully determined by its representational properties what results is the following thesis:

(KREP) A thought or experience R presents the world as seeming a certain way when a subject can have first-person knowledge of R's representational properties.

(KREP) claims we can know what it is like to undergo an experience from the inside, and this is surely right: we do for the most part know what it is like to undergo our own experiences just by having them. It strikes us as far-fetched to raise doubts about my sincere claim that "I am currently in pain".⁵ This, I take it, is because we have a warrant

⁵ Wittgenstein (1953: 246) famously goes as far as to claim that it is for this reason a mistake

Consciousness, Self-Consciousness and First-Person Knowledge

for such assertions that derives from the consciousness that attaches to them. To doubt a sincere utterance of “I am currently in pain” is to question whether a person does indeed have warrant for making such an utterance. This is something it doesn’t occur to us to do because a person that says sincerely that they are in pain is a person that is conscious of being in pain. Just by consciously experiencing pain such a person has warrant for saying they are in pain.

(KREP) doesn’t attribute infallibility to a subject: it allows that a subject may well be mistaken when he says he is feeling something painful. The subject may for instance be hypnotised so that he believes that lemons taste sweet. We wouldn’t say that such a subject knows that the lemon he has just eaten tastes sweet to him, since beliefs formed as a result of hypnotism have not been formed in the right way to count as knowledge. (KREP) doesn’t require us to say that this subject knows that when he tastes a lemon, the experience he has just had is of something that tastes sweet. (KREP) claims only that a subject *could* know this in a direct and immediate way. Saying that a subject could have first-person knowledge of his experiences is quite consistent with allowing that a subject may on occasion fail to achieve such knowledge.

One difficulty for (KREP) is that it seems to conflict with the transparency of experience which the intentionalist is so keen to stress.⁶ It suggests that when a subject

to talk of a subject’s *knowing* he is in pain at all. We talk of knowledge only where there is room for doubt, but since there is little or no room for doubt in cases like this one, it makes no sense to talk of a subject’s knowing he is in pain. If Wittgenstein is right (KREP) looks misconceived: (KREP) talks of us having knowledge of our experiences but Wittgenstein tells us such talk is at best misplaced.

Utterances of the form “I know I am experiencing x” seem to me to be meaningful and to carry information that isn’t carried simply by saying “I am experiencing x”. I take our talk of first-person knowledge to be significant because it communicates something of the distinctive kind of warrant that attaches to first-person knowledge as contrasted with third-person knowledge – knowledge of another person’s experiences. For a further discussion of Wittgenstein on first-person knowledge with which I am in full agreement see Siewert (1998: pp.27-33).

⁶ For the use that intentionalists make of transparency, see Harman (1990/1997) and Tye (1992). Martin (2003) uses transparency type considerations to raise difficulties for an intentionalist account of conscious experience. It should be noted that Tye isn’t as careful as he might have been in his discussion of transparency. In a much quoted passage Tye tells us that when he found himself transfixed by the blue of the Pacific Ocean what he was

Consciousness, Self-Consciousness and First-Person Knowledge

gains introspective knowledge of her experiences she comes to know the representational properties of her experience. The intentionalist will object that when one introspects, one doesn't find anything like an experience and its representational properties. Instead one finds items belonging to the public world and their properties. This suggests that the knowledge we gain of our experiences through introspection isn't knowledge of an experience's representational properties, but is rather knowledge of whatever is represented.⁷

I take it that accommodating something like this worry is the motivation for Dretske's (1995: ch.2) accounts of introspection as "displaced perception". He says that we come to know what an experience is like through the awareness we have the objects of our experience in much the same way as we come to know of our weight by means of bathroom scales.⁸ We do not come to know what an experience is like by coming to know of its representational properties, but only by coming to know what an experience represents.

If (KREP) is to be made compatible with transparency we must modify its formulation to allow that what a subject knows in the first-person isn't an experience's

enjoying was "an aspect of the content of his experience" (*ibid*, 160). If our experiences have the contents they do in virtue of their representational properties, it follows that what Tye was enjoying was the representational properties of his experience.

I shall argue that if we are to do full justice to the claim that experiences are transparent we must think of introspection as making us aware of whatever it is our experience represents. Dretske offers an account of introspection which acknowledges this point. I take Dretske to be committed to denying that "what is immediately accessible to consciousness" to borrow Tye's words, is an experience's content, though I have not managed to find anywhere where he says as much. In any case, I shall say that it is not an experience's representational properties that are immediately accessible to me when I introspect. An experience's representational properties are properties of the experience *qua* representational vehicle, and I have no access to those properties. Rather what is made accessible to me when I introspect is whatever it is, if anything, that my experience represents at the time. This will mean revising the formulation of KREP somewhat as we shall see in due course.

⁷ This characterisation of introspective knowledge doesn't require us to deny that there is any difference between first-person and third-person knowledge. It can allow that we know our own experiences directly and immediately whereas we know the experiences of others only via observations. For an account of introspection along the line just sketched which nevertheless recognises the distinction between first-person and third-person knowledge, see Evans (1982, ch.7, §7.4) and Dretske (1995, ch.2, pp.51-4).

⁸ Dretske (1995: 41)

Consciousness, Self-Consciousness and First-Person Knowledge

representational properties, but is rather whatever the experience represents by means of those properties. I propose then that we revise (KREP) as follows:

Consciousness, Self-Consciousness and First-Person Knowledge

(KREP*): An experience E seems or feels a certain way iff a subject can have first-person knowledge of whatever E represents.

We should now have a better sense of the question this chapter will address and the strategy I will employ for answering it. The question asks: what is the difference between representational states that make a subject conscious of something and those that do not? I have said that a representational state makes a subject conscious of something – it presents the world as seeming or feeling some way – when it can be known from the inside just by being in it. My strategy for answering our question is to establish just how it is that the representational states which make us conscious of something can be known from the inside. How is it that a subject can know he is in a particular representational state just by being in that representational state? Answering this question will tell us what it is for a representational state to have phenomenal character.

In the next section we shall see that there are two possible answers to this question. The first appeals to something extrinsic to a representational state to explain how it gets its phenomenal character. The second answer makes phenomenal character something intrinsic to a representational state. The first task we face then will be to decide between these two possibilities.

2. Two Conceptions of Phenomenal Character

(KREP) in both of its formulations is consistent with two distinct ways of thinking about an experience's phenomenal character.⁹ We can think of phenomenal character as something *of* which the subject is consciously aware. Alternatively we can think of it as

⁹ I am indebted here to Rowlands (2001, ch.6). Rowlands notes that "Consciousness can be both object and act of experience. Metaphorically speaking, consciousness can be both the directing of awareness and that upon which experience is directed. Consciousness can include both experiential features of which we are aware, and experiential features with which we are aware." (*op cit*, 122) While Rowlands recognises that there are these two distinct ways of conceiving of consciousness he argues against what he calls an "objectualist" conception of phenomenal character as something of which the subject is aware.

Consciousness, Self-Consciousness and First-Person Knowledge

something *with* which a subject is consciously aware of the world. Any answer to our question of why there is something rather than nothing an experience is like must decide between these two conceptions of phenomenal character.

To get a better fix on these two conceptions consider by way of illustration the experience one has when tasting honey. The first conception of phenomenal character has it that the sweet taste is something of which the subject is conscious. The sweetness the subject experiences is itself something *of* which the subject is conscious.¹⁰ On the second conception of phenomenal character, the sweetness is something the subject experiences in virtue of his sensory experience's intentional content. As such it is something *with* which the subject is conscious of the honey. The conscious experience is understood as that by means of which a subject's consciousness is directed towards a particular object, in this case the honey.

The first conception of phenomenal character is naturally combined with higher-order theories of consciousness. Higher-order theories claim that a mental state has phenomenal character only if a subject is conscious of being in that mental state. A perceptual experience won't seem or feel any way to a subject unless the subject is conscious of having this experience. A perceptual experience might well represent a creature's local environment as being a certain way; it may for instance represent some object as coloured. The experience the subject is having of a coloured object won't seem a certain way to the creature unless it is conscious of having this experience. On this account of consciousness, phenomenal character makes a subject conscious of things in her environment only if she is first conscious of having an experience of those things. It is

¹⁰ A proponent of the first conception of phenomenology isn't committed to the existence of irreducibly phenomenal properties of the kind discussed in chapter 3. He could say that when a subject is conscious of an experience's phenomenology he brings to bear certain recognitional capacities. He might be recognising for instance the property in virtue of which something is experienced as tasting sweet. On this view an experience seems or feels a certain way by being made available to certain recognitional capacities in virtue of which a creature can discriminate among an experience's representational contents. For an account along these lines see Carruthers (2001).

Consciousness, Self-Consciousness and First-Person Knowledge

only if a subject is conscious of having this experience that she will enjoy an experience that seems or feels a certain way.

The second conception fits with the intentionalist account of consciousness, according to which phenomenal character is something *with* which the subject is conscious of things in its environment, but is not something *of* which the subject is conscious. The intentionalist claims that conscious experiences are nothing but vehicles for making a subject conscious of things in its local environment. The experience isn't itself something of which the subject is conscious until the subject turns his gaze inward, and in the process makes an experience the explicit object of his attention. Otherwise his attention is with whatever he is experiencing.

However the intentionalist still owes us an account of how a representational state can be known from the inside. Tye (1995) gives us what looks like an answer to this question. He tells us that an experience has phenomenal character when its content is Poised Abstract Nonconceptual Intentional Content, which he abbreviates as PANIC. Representations that differ in their PANICs differ in their phenomenal character. A representation has a content that is poised when it 'stands ready and in position to make an impact on the belief/desire system.' (Tye, 1995: 138)

However to account for knowability from the inside in terms of poise is to presuppose what one sets out to explain. If am right an experience exhibits what Tye calls "poise" only because it has phenomenal character. Appealing to poise cannot explain what it is for a state to possess phenomenal character. Poise has built into it the property it is supposed to explain. At least this will follow if we accept (KREP*) – the claim that what it is for an experience to present the world as seeming a certain way is for that experience to be knowable from the inside.

Higher-order theories are designed to answer the question of what makes a state knowable from the inside. They say a state is like something for a subject when a creature is either disposed to produce, or actually produces, a higher-order representation

Consciousness, Self-Consciousness and First-Person Knowledge

of a thought or experience he is having. In this respect, higher-order theories of consciousness look to have an advantage over intentionalism. We shall see however that this advantage is only superficial. Although Tye's account is unsatisfactory, I will suggest that the phenomenologists succeed where he fails. Thus it will turn out that we have two rival accounts of what it is for a state to have phenomenal character.

The story begins with Brentano. Following the famous chapter on intentionality in his *Psychology from an Empirical Standpoint*, Brentano set about defining what it is for a thought or experience to be conscious.¹¹ Our conscious thoughts and experience he argues are directed towards two objects. First we are conscious of what Brentano calls a "primary object". We can call this kind of consciousness "outer consciousness" since it is consciousness that is directed outwards to something extra-mental in the world. The primary object is the intentional object which a thought or experience is directed towards. Drawing on Husserl I suggested in the previous chapter that we think of the intentional object as an object the thought or experience purports to represent, where "object" is to be construed in a broad sense as including a thing, event, process, condition or state of affairs.

Brentano claims that in addition to being directed towards a primary object conscious experiences can also be directed upon themselves. Consider the experience of hearing a sound. The sound is the primary object, and the experience of hearing the sound is a secondary object. An experience represents a primary object (the sound) by purporting to refer to a sound originating from something in the subject's local environment. At the same time it represents a secondary object, its own occurrence, by being directed upon itself. The consciousness a subject has of an experience she is having we can call "inner consciousness". Brentano's claim is that a single representational state can instantiate both inner and outer consciousness.

¹¹ Brentano's own view is introduced in §7 of Chapter 2 entitled 'A Presentation and the Presentation of that Presentation are Given in One and the Same Act.'

Consciousness, Self-Consciousness and First-Person Knowledge

Does this Brentanian account of consciousness supply the intentionalist with an account of phenomenal character? Let us consider first how Brentano might have understood phenomenal character. When an experience represents its own occurrence, a subject thereby has epistemic access to this experience from the inside; she can know that she is the subject of this experience just by having it. I said earlier that a subject has this kind of epistemic access only if her experiences manifest a phenomenal character. Brentano might have concluded then that an experience has a phenomenal character only when it represents its own occurrence. An experience that doesn't represent its own occurrence won't seem or feel anyway to the subject undergoing this experience.¹²

Consider by way of illustration, my experience of hearing a blast of sound from a trumpet. A Brentanian account of consciousness would claim that at the same time as I hear the trumpet my act of hearing is directed upon itself. By being directed upon itself my experience makes me conscious of its occurrence. Brentano might have said that it is the consciousness I have of an experience's occurrence that makes an experience's occurrence like something for me. When I am not conscious of my experience's occurrence, I can hear a trumpet without it seeming or feeling any particular way to me.

Does an account of phenomenal character along these lines provide the intentionalist with an answer to the question of when a representational state instantiates a phenomenal character? Recall that the intentionalist conceives of phenomenology as something with which, but not of which the subject is conscious. Is the Brentanian view of phenomenal character just sketched consistent with intentionalism?

Suppose we think of phenomenal character as something of which a subject is conscious. Then it will be natural to think that it is only by means of an additional representational state that an experience could possibly have phenomenal character. For it is only by means of an additional representational state that a subject could be conscious of an experience's phenomenal character. Brentano denies that a subject's

¹² For a recent defence of such a view see Kriegel (forthcoming)

Consciousness, Self-Consciousness and First-Person Knowledge

experience needs to be accompanied by a higher-order representation in order to have a phenomenal character. He claims that a single representational state can be directed at something in a creature's environment, and at the same time also be directed upon itself. It might be thought that because Brentano rejects a higher-order account of phenomenal character he must instead hold an intentionalist view of phenomenal character.

A moment's reflection will suffice to show that such a conclusion is mistaken. The Brentanian account of phenomenal character sketched above is in important respects different from that proposed by higher-order theories,¹³ but Brentano shares with these theories a view of phenomenal character as something *of* which a subject is conscious. Brentano's theory of consciousness and higher-order theories both agree that a representational state has phenomenal character only if a subject is conscious of being in that representational state. Brentano claims that when an experience represents its own occurrence the result is that the subject becomes conscious *of* having this experience: a subject becomes conscious of the experience as a secondary object. The experience becomes an object (a secondary object) for the subject in addition to whatever extramental object the experience is representing at the time. Thus Brentano would appear to be committed to the claim that (*pace* intentionalism) a subject is conscious of his experience's phenomenal character.

The intentionalist claims that conscious experiences are something of which a subject is conscious only when the subject engages in an act of introspection. Otherwise the

¹³ Higher-order theories conceive of consciousness as a relational property which is instantiated when an appropriate relation holds between a first-order representational state and a higher-order representation. Consciousness is conceived of as a higher-order monitoring of lower level states and processes. Brentano is committed to a one-level account of consciousness which makes consciousness into an intrinsic property of at least some mental states. According to his account a mental state is conscious when it is directed upon itself at the same time as it is directed at something in the world: a representational state needn't actually or potentially stand in a relation to any other mental state in order to instantiate consciousness and phenomenal character. It should be noted that Brentano argued that consciousness was a property of *all* mental states. In other words he denied that there are any unconscious mental states. Kriegel (2003, pp114-116) argues that Brentano might have said all *occurrent* mental states are conscious, making the claim less susceptible to obvious counterexamples.

Consciousness, Self-Consciousness and First-Person Knowledge

subject is conscious of whatever his experience represents. Thus the intentionalist wants to combine the following seemingly incompatible theses:

(1) Mental states that present the world as seeming a certain way can be known in the first person.

(2) It is possible for a mental state *M* to be conscious, although a creature is not conscious *of* itself being in *M*.

Theses (1) and (2) seem to conflict; a mental state that can be known in the first-person must be one to which the subject has epistemic access; it must somehow present itself to a subject from the inside. How can a subject have epistemic access to a mental state without the subject also being conscious of itself in this mental state?

I will suggest in the next section that phenomenologists may have found a solution to this problem, thereby supplying the intentionalist with an account of when a representational state can be said to have phenomenal character. This will still leave us needing to choose between the two accounts of phenomenal character introduced in this section. In section 4 I will offer further arguments against the first conception of phenomenal character by arguing against higher-order and Brentanian theories of consciousness both of which subscribe to this view of phenomenal character.

3. Consciousness and Self-Consciousness

Phenomenologists claim that every conscious experience has built into it a variety of self-consciousness. We saw in chapter 4 that the phenomenologist claims our experiences are intentional before a subject takes up any relation to the world. The phenomenologist says something analogous about being self-conscious.¹⁴ He denies that a subject's

¹⁴ I will support my claims primarily by reference to Sartre, but it should be noted that there are similar ideas to be found in Husserl as Zahavi (1999) has persuasively argued. I will discuss the views of the other phenomenologists in chapter 7 where I develop a naturalistic account of phenomenal character.

Consciousness, Self-Consciousness and First-Person Knowledge

experiences are made self-conscious by a subject reflecting on his experiences, but equally he will reject a theory along Brentanian lines which says that our experiences are made self-conscious when a mental state is directed upon itself. More generally this theory denies that self-consciousness can be understood as any kind of relation between a subject and her mental states. If it is a mistake to conceive of self-consciousness as a relation between a subject and her mental states, it is equally a mistake to view self-consciousness as a relation between a representational state and itself. Our experiences are already self-conscious before the subject takes up any relation to them.¹⁵ Sartre has explained the idea as follows: “This self-consciousness we ought to consider not as a new consciousness, but as the only mode of existence which is possible for a consciousness of something.” (Sartre, 1943/2000: xxx)

The phenomenologist makes an important distinction between pre-reflective and reflective *self*-consciousness. It is reflective self-consciousness that philosophers commonly discuss when they are concerned with self-consciousness.¹⁶ Reflective self-consciousness is the consciousness a subject has of herself through introspective reflection. When a subject introspects she forms a belief that she is herself in a mental state M. I will have more to say about what this involves in a moment. Suffice it to say that the phenomenologist is *not* claiming that conscious experiences have built into them reflective self-consciousness. He claims that a subject can be conscious *of* herself as being in a mental state M only because M is already accompanied by a kind of self-consciousness. I shall call the kind of self-consciousness in question “pre-reflective self-consciousness”. It is pre-reflective self-consciousness which it is claimed forms a part of every conscious experience. Before I consider the phenomenologist’s argument for this

¹⁵ This doesn’t mean that the phenomenologist thinks of self-consciousness as primitive and unanalysable: we shall see in section 5 that the phenomenologists have plenty to say about what it is for a mental state to be intrinsically self-conscious.

¹⁶ Recently pre-reflective self-consciousness has received some attention under the heading of non-conceptual self-consciousness, see for instance Bermúdez (1998: ch.5-8) and Hurley (1998: ch.4).

Consciousness, Self-Consciousness and First-Person Knowledge

claim we need to get a firmer grip on the difference between reflective and pre-reflective self-consciousness. This will be my aim in the remainder of this section.

Reflective self-consciousness attaches to the judgement a subject makes when she self-ascribes some thought or experience she is having. When a subject self-ascribes a conscious state or activity M she judges that she herself is in a mental state M. I have spoken of the subject judging that “she herself” is in M to capture the fact that the judgement the subject makes is one she would express in the first-person. It is a judgement which if I were making I would express by saying “I am in M”. I will follow Castañeda by putting an asterisk beside a pronoun to signal cases of first-person reference. Thus I will henceforth abbreviate “she herself” as “she*” and “he himself” as “he*”. When a subject self-ascribes a mental state M she judges that she* is in M.

To see that a subject must be reflectively self-conscious when a subject judges that she* is in M, notice that the content of S’s judgement is not <I am S and S is in M>. When S self-ascribes a mental state M she doesn’t need to first identify an individual S and then establish that it is S that is in M.¹⁷ S doesn’t need to take this complicated route because when she thinks about herself in the first-person (as she does when she judges that she* is in M) she cannot fail to realise that she is thinking about herself.¹⁸ It doesn’t make sense for her to think that ‘she* is in M’ and to also think ‘someone is in M but is it myself that is in M?’¹⁹ This is because when a subject thinks of herself by means of

¹⁷ To borrow terminology from Evans (1982), S’s judgement that she* is in M is “identification free”. Evans applies the term to knowledge that is arrived at in a certain way but I do not think I am distorting the concept by taking it to apply to judgements. According to Evans knowledge is identification free if “(1) it is not identification dependent and (2) it is based on a way of gaining information from objects” (Evans, 1982: 181). While knowledge of a proposition that ‘x is M’ is *identification dependent* if it is inferred from the propositions that ‘y is M’ and ‘x = y’. Knowledge is identification free then, if the subject can know that x is M without first establishing that ‘y is M’ and ‘x = y’.

¹⁸ To use an example from Castañeda (1966), there is a difference between thinking that the Editor of Soul is a millionaire and thinking I am a millionaire even when one is the Editor of Soul. One can fail to know that when one thinks of the Editor of Soul one is thinking about oneself, but this is not true when one thinks of oneself by means of “I”. One cannot fail to realise that it is oneself that one is thinking of when one entertains a thought in the first-person. I will suggest that this is because such thoughts are had self-consciously.

¹⁹ Shoemaker (1968) describes this semantic phenomenon as “immunity to error through misidentification relative to the first person pronoun” (“IEM”). I shall follow Shoemaker in

Consciousness, Self-Consciousness and First-Person Knowledge

“I” she is self-conscious. The kind of self-consciousness that accompanies first-person thoughts of this kind (thoughts the subject has in the course of self-ascribing a mental or bodily state) is what I am calling “reflective self-consciousness”.

Let us henceforth conceive of “reflective self-consciousness” as the self-consciousness that accompanies a person’s thought that she* is in a mental state M, in virtue of which a person cannot fail to realise that she is thinking about herself. I shall henceforth call thoughts that are reflectively self-conscious “I-thoughts”. Modifying Rosenthal’s (1990/1997) terminology somewhat we can characterise reflective self-consciousness as a transitive form of *self*-consciousness. When a subject makes a judgement that she* is having an experience M she predicates a property of herself, the property of having M. Reflective self-consciousness is consciousness that one is oneself the subject to whom this predicate pertains.

I have said above that the phenomenologists deny that pre-reflective self-consciousness is a relation between a subject and some mental state she is in at the time. They deny then that pre-reflective self-consciousness is a transitive form of self-consciousness. If pre-reflective self-consciousness isn’t a transitive form of self-consciousness might we instead characterise it as an *intransitive* form of self-consciousness?

Kriegel (2004) argues for the existence of something he calls “intransitive state *self*-consciousness”. Rosenthal (1990/1997) introduced the notion of “intransitive *state* consciousness” to characterise our use of the term “conscious” when we say that a token thought or experience has the property of being a *conscious* thought or experience.²⁰ Kriegel argues that a subject can instantiate this property only if she is in a mental state that is intransitive state *self*-conscious. Substitute “intransitive state self-consciousness”

taking IEM to be a feature peculiar to self-consciousness. I shall return to IEM in more detail in section 4.

²⁰ Rosenthal goes on to argue that it is a relational property which is instantiated when a mental state is simultaneously accompanied by a higher-order thought. I will have more to say about this later in section 4.

Consciousness, Self-Consciousness and First-Person Knowledge

for “pre-reflective self-consciousness”, and we get the thesis I have attributed to the phenomenologists that a mental state can be conscious only if it is pre-reflectively self-conscious. Let us consider then whether Kriegel’s notion of “intransitive state self-consciousness” might give us an initial handle on pre-reflective self-consciousness.

Kriegel introduces the notion of intransitive state self-consciousness by asking us to consider the difference between the following two sentences attributing self-consciousness to a subject:

- (1) Smith is self-conscious of thinking that her car is new.
- (2) Smith is self-consciously thinking that her car is new.

The first sentence attributes to Smith what I have called “reflective self-consciousness”. I have argued that a subject can be reflective self-conscious only by being in a distinct mental state M_2 which makes her conscious of herself being in a mental state M_1 . Thus (1) describes a case of transitive creature self-consciousness. The second sentence attributes to a subject intransitive state self-consciousness. The self-consciousness which (2) reports, isn’t brought about through a distinct act of reflection. The state in virtue of which Smith is self-conscious is one and the same as the state which is her thinking about her new car. We can say that the thought about her new car is a first-order property of Smith, and the self-consciousness is a second-order property of Smith which “modifies”, to borrow Kriegel’s term, Smith’s thinking.

What is it to think or experience self-consciously? It is to be conscious of oneself having a thought or experience. This is not to say that one is conscious of oneself having a thought or experience in the same way as one is conscious of the object of one’s thought or experience. The self-consciousness that (2) attributes isn’t the result of a distinct act of reflection whereby a subject makes a mental state she is undergoing into the object of her thought. Rather the state in virtue of which Smith is self-conscious is the same state that makes her conscious of her new car. Smith’s being self-conscious is a matter of the

Consciousness, Self-Consciousness and First-Person Knowledge

way in which *she* has this thought. Her self-consciousness modifies the way in which this thought occurs to her, distinguishing it from a thought that is not had self-consciously.

Kriegel has suggested that the kind of self-consciousness (2) describes is best thought of as an “implicit” or “peripheral” consciousness one has of oneself having a conscious thought or experience. When one undergoes a conscious experience one is aware of oneself having this experience. To be aware of oneself having an experience is to be aware of oneself *qua* subject of experience. It is subjects that have or own experiences. The awareness one has of oneself *qua* subject isn’t explicit. It would only become explicit if one could somehow token a thought which targeted the awareness one has of oneself having an experience.²¹ We have already seen that intransitive state self-consciousness involves no such act of reflection. Instead the very same experience one is undergoing is such that one is conscious of oneself having it at the same time as one is made conscious of its object. It is this implicit or peripheral consciousness one has of oneself *qua* subject of a thought or experience which Kriegel calls “intransitive state self-consciousness”.

Did the phenomenologists have something similar in mind when they talked of pre-reflective self-consciousness? Sartre will often talk of “consciousness being conscious of itself as consciousness of an object.”²² Where Sartre talks of “consciousness” I shall instead talk of “subjects of consciousness”. The proposal I am considering takes Sartre to be saying that when a subject is conscious of an object there are two properties instantiated. The first is the property in virtue of which the subject is conscious of an object. I shall call this property “P₁”. The second is the property in virtue of which the

²¹ It is far from clear that this is something we can do. When Hume worries (to paraphrase) that he can not catch himself without a perception and never can observe anything but perception, it might be thought that what he has noticed is that one can never become aware of himself as the subject of his experiences. As soon as he tries to observe himself having an experience, he becomes the person doing the observing and all he discovers are the mental states that he is undergoing. What Hume cannot observe is himself as the subject that is undergoing those mental states. I will return to this worry at the end of section 5.

²² This is a formulation that recurs repeatedly in his (1943/2000), see for instance Section 3 of the Introduction titled ‘The Pre-Reflective Cogito and the Being of the Percipere’.

Consciousness, Self-Consciousness and First-Person Knowledge

subject is conscious of itself as conscious of an object. Call this second property “P₂”. The proposal I am considering characterises “P₂” as a second-order property, a property of P₁. P₂ characterises the way in which a subject is conscious of an object. It is a property of P₁ which modifies a subject’s consciousness of an object, making the subject conscious of himself having a particular thought or experience.

When Sartre says the subject is conscious of himself, he is not saying the subject is himself among the objects he experiences when he has a conscious experience. In fact it is central to Sartre’s philosophy to deny that the subject is ever aware of himself as an object. For Sartre consciousness is empty: it is always directed towards something that is not itself. We find here an echo of the intentionalist’s claim that phenomenal character is that with which, but not of which the subject is conscious. Sartre generalises this conclusion to claim that there is literally nothing in the mind of which we are conscious. Even when a subject engages in reflection a subject’s thinking is directed towards something that is not itself. Introspective reflection issues in knowledge, and knowledge is characterised by a duality of knower and that which is known. When the subject knows itself through introspection “the reflected on must necessarily be the object for the reflected; and this necessarily involves a separation of being” (Sartre, 2000: 151).²³

In what sense then is a subject conscious of *herself* when she has an experience self-consciously? When one perceives an object one is able to identify that object, singling it out from other objects with which one is simultaneously presented. I take Sartre (and indeed phenomenologists more generally) to be committed to the view that we can single out an object only if we know the object’s location in relation to oneself. To know where an object is in relation to oneself one must be conscious of the position of one’s body in

²³ I take Sartre’s point to be similar to one that Ryle (1949) makes when he notes that trying to make oneself into the object of one’s thought is like trying to jump on one’s own shadow. Every time you make the leap your shadow moves away from you, and the same is true of the self. Every time I try to observe myself by engaging in an act of reflection, I myself move out of view as the subject that is doing the observing. Sartre’s point seems to be that no person can simultaneously occupy the position of observer and that which is observed.

Consciousness, Self-Consciousness and First-Person Knowledge

relation to this object. I also take Sartre to be committed to the view that when I identify an object's position in relation to my body I do not have to identify a body as my own body. The knowledge I have of my own position in space is "identification free" to borrow a term from Evans (1982).²⁴ I know directly and immediately without having to form any observations of my body its location relative to the object that I perceive. This is because at the same time as I am conscious of an object I am conscious of myself. The consciousness I have of myself is also a consciousness of my body and the way in which the world is presented to it. This consciousness makes it possible for me to know where an object is in relation to me without my first needing to identify my own location.²⁵

What I have said so far suffices for us to see that the suggestion from Kriegel that we view pre-reflective self-consciousness as an implicit or peripheral consciousness of self seems to be along the right lines. It is the implicit consciousness a subject has of herself as the subject of an experience which enables her to identify the location of objects relative to her own position in space without first having to identify her own location in space.

I suggest we construe the distinction between reflective and pre-reflective self-consciousness as follows. I have endorsed the intentionalist view that to be conscious is to be in a certain kind of representational state. Expanding on this proposal I suggest we draw the following distinction between types of conscious representational state.

- (1) A subject is reflectively self-conscious when she *explicitly* represents that she is the subject of some thought or experience.

²⁴ See footnote 19 for a definition of identification free knowledge.

²⁵ I discuss this line of reasoning in more detail later in chapter 7.

Consciousness, Self-Consciousness and First-Person Knowledge

(2) A subject is pre-reflectively self-conscious when she *implicitly* represents to herself that she is the subject of some bodily or mental state.

A subject *explicitly* represents herself when she represents herself *as* the subject undergoing a conscious state or activity. She explicitly represents herself only by deliberately attending to her conscious states thereby thinking of herself as the subject of some conscious state or activity. What is it for a subject to *implicitly* represent herself? This is a question which will be discussed briefly in chapters 6 & 7, but we are already in a position however to sketch the beginning of an answer to this question. First of all when a subject is implicitly represented she does not figure among the things she represents. I said above that we should understand pre-reflective self-consciousness as *modifying* the consciousness a subject has of an object. I suggested we think of a subject's being conscious of an object as a first-order property of the subject, and pre-reflective self-consciousness as a second-order property of the subject. Thus implicit representation is a property which somehow modifies a representational state which makes a subject conscious of an object.

One possible way to think about a subject's implicitly representing herself is to say that a conscious state isn't just a representation of an object but is a representation of the object as standing in a relation to the subject. I have said that a representational state is somehow modified when it implicitly represents the subject that is in it. The current suggestion is that it is modified in such a way that a subject represents an object as standing in a relation to herself. By representing an object in a relation to herself she is implicitly aware of herself at the same time as she is explicitly aware of the object. Thus we can say that every conscious state has a content of the following form $R\langle I, f \rangle$, where 'I' stands for the subject of this conscious state, 'f' for an object the state represents and R for a relation the state represents as holding between I and f.

How does this distinction between reflective and pre-reflective self-consciousness help

Consciousness, Self-Consciousness and First-Person Knowledge

us with our question of what it is for an experience to have phenomenal character? I have said that the phenomenologist takes reflective self-consciousness to depend upon pre-reflective self-consciousness. This is to say that one can become reflectively self-conscious only because one is already pre-reflectively self-conscious. On this theory a subject can judge that she* is in a mental state M only because the mental state M she self-ascribes is already pre-reflectively self-conscious. The judgement the subject makes in self-ascribing a mental state will normally yield first-person knowledge. I have said in section 1 that a subject can have first-person knowledge of a mental state only if that mental state has phenomenal character – if it seems or feels a certain way to its subject. We have just seen that the phenomenologist holds a mental state can be known in the first-person only if it is already pre-reflectively self-conscious. I take this thesis to form the basis for a proposal about what it is for an experience to have phenomenal character. According to this proposal an experience has phenomenal character only if it is pre-reflectively self-conscious. It is only if the subject implicitly represents herself having an experience that the experience she has will seem or feel a certain way.

Such a proposal is consistent with intentionalism: it allows that phenomenal character is something with which we are conscious but not something of which we are conscious. For recall that pre-reflective self-consciousness is not transitive self-consciousness: when an experience is had self-consciously the subject is not explicitly conscious of herself, she is not conscious of herself *as* the subject having an experience. The subject only becomes explicitly conscious of herself as the subject of an experience when she makes an experience the object of her attention, through introspection. Recall that this is just what the intentionalist claims about phenomenal character in defending the idea that conscious experiences are not experiences of which the subject is conscious.

The claim that conscious states are pre-reflectively self-conscious explains how a conscious state can seem or feel a certain way without the subject being conscious of this state. The state can seem or feel a certain way because it is had by a subject in such a

Consciousness, Self-Consciousness and First-Person Knowledge

way that the subject is conscious that she stands in a relation to whatever the state represents. When the subject represents an object as standing in a relation to herself, the state needn't figure among the things of which she is conscious. She can be aware of nothing but the object she represents so long as she is aware of this object as standing in a relation to herself. By representing an object in relation to herself, she becomes aware of being in a particular representational state. She is aware of herself as being conscious of an object, but she is aware of being in this state without actually representing that she is in this state.

We now have two genuinely alternative proposals corresponding to our two conceptions of phenomenal character.

Proposal 1: An experience E seems or feels a certain way to me iff E is accompanied by a higher-order representation that I am having E.

Proposal 2: An experience E seems or feels a certain way to me iff E is pre-reflectively self-conscious.

In the next section we will see that to defend proposal 2 over proposal 1 it must be shown that reflective self-consciousness depends on pre-reflective self-consciousness. When the phenomenologist claims that reflective self-consciousness depends on pre-reflective self-consciousness she is saying that a subject couldn't be reflectively self-conscious – she couldn't self-ascribe a mental or bodily state of hers – unless the mental state she was self-ascribing was itself already pre-reflectively self-conscious. I shall call this “the dependence thesis”. The dependence thesis has been asked to do a lot of work in the account of phenomenal character I have just attributed to the phenomenologist. I have said that a state is knowable in the first person when it has phenomenal character. The dependence thesis says that a state is knowable in the first person only if it is pre-reflectively conscious. From these two theses I derived the phenomenologist's account of

Consciousness, Self-Consciousness and First-Person Knowledge

phenomenal character which says that a state has phenomenal character only if it is had self-consciously (that is to say only if the mental state is pre-reflectively self-conscious).

Proponents of a higher-order theory of consciousness would reject the dependence thesis. They would claim that a subject becomes self-conscious through an act of reflection. Thus a subject doesn't need to already be pre-reflectively self-conscious in order to achieve reflective self-consciousness. If the dependence thesis can be defended this will serve to motivate the phenomenologist's account of phenomenal character over the higher-order theorist (and the Brentanian for that matter). We will have shown that we can account for reflective self-consciousness only by granting that the subject is already pre-reflectively self-conscious. This amounts to conceding that an experience can only be known in the first person if it is pre-reflectively self-conscious. Recall that this is just what the phenomenologist claims when she says that an intentional state has phenomenal character only if it is pre-reflectively self-conscious. Thus by defending the dependence thesis we get an argument for the phenomenologist's account of phenomenal character. The next section will consider how phenomenologists argue for the dependence thesis.

4. The Dependence Thesis

I shall begin by sketching an outline of an argument for the dependence thesis. The argument proceeds by attempting to show that reflection by a subject on his mental states couldn't possibly make a subject self-conscious.²⁶ By "reflection" I mean the ability a subject has to introspect, where introspection can be understood either as a form of perception or "internal sense" as it was by Locke or as thought about one's own mental

²⁶ I must stress that the following argument doesn't appear anywhere in the phenomenologist's writing, so far as I am aware. I am constructing this argument on the phenomenologist's behalf based on the priority that Sartre and Merleau-Ponty give to the notion of pre-reflective self-consciousness and the treatment these philosophers have been given in secondary literature. See in particular the excellent studies by Wider (1997) and Zahavi (1999).

Consciousness, Self-Consciousness and First-Person Knowledge

states as it is by Rosenthal (2004).²⁷ I shall call theories which take self-consciousness to have its origins in reflection “reflection theories”. The phenomenologist will argue that self-consciousness cannot originate from a subject observing or thinking about his own mental states. Yet we have seen that when a subject introspects and self-ascribes a mental state he thinks about himself self-consciously. Since this self-consciousness isn’t brought about through an act of reflection, the phenomenologist infers that the subject must already be in a mental state that is self-conscious before he introspects. It is this self-consciousness that they call “pre-reflective self-consciousness”.

There are a number of points at which this sketch of an argument can be challenged. First the phenomenologist must show that reflection cannot yield self-consciousness. This means showing that self-consciousness cannot come about either by a subject observing her own mental states or by a subject thinking about her own mental states. We shall see that higher-order perception theories challenge the first disjunct: they might claim that a subject can become self-conscious by perceiving her conscious mental states.²⁸ Higher-order thought theories challenge the second disjunct: they claim that a subject can become self-conscious by thinking about her own conscious mental states.

Let us suppose that the phenomenologist can justify the claim that self-consciousness doesn’t have its origins in reflection. Still someone might be unwilling to concede to the phenomenologist, the conclusion that our mental states are intrinsically self-consciousness. He might worry that the phenomenologist’s argument assumes that the only two options are to claim that self-consciousness comes about through reflection or to accept that some mental states are intrinsically self-conscious, but why suppose that these two options are jointly exhaustive of the possible positions one might take on this

²⁷ These are two very different conceptions of introspection but I won’t attempt to decide between them at this stage. We shall see that the first conception of introspection is defended by higher-order perception theorists and the second conception of introspection is defended by higher-order thought theories. Both are not without problems as we shall see.

²⁸ I confess that I have been unable to find such a theory explicitly asserted by proponents of higher-order perception theories (e.g. Lycan and Armstrong). I have arrived at such a theory based on the account they give of introspective knowledge.

Consciousness, Self-Consciousness and First-Person Knowledge

question?

This second objection is the easiest to deal with, so I shall start my defence here. First of all we should note that the phenomenologist is treating the question of how a subject comes to be self-conscious as a question that is to be given a personal-level answer. By a “personal level explanation” I mean an explanation that makes ineliminable reference to persons and their mental states.²⁹ Of course there is another possibility the phenomenologist doesn’t consider which is that the question can be answered by appeal to theories that invokes *sub*-personal mechanisms. We have seen earlier that the phenomenologist rejects the possibility of explaining consciousness in sub-personal terms. I shall return to the argument they give for this conclusion towards the end of this section. In barest outline the worry they have is that such an explanation of self-consciousness will leave something out. It will treat the subject as one material thing among others, but in the process fail to account for the sense we have of ourselves as *subjects* when we are pre-reflectively self-conscious. We have already encountered this objection in chapters 2 and 3, but we will see that it can be made more precise by relating it to the phenomenon of pre-reflective self-consciousness.

Let us set aside for now, the objection that the phenomenologist has ignored the possibility of giving a sub-personal explanation of self-consciousness. Is the phenomenologist nevertheless right to think that the only possible *personal* level explanations treat self-consciousness either as the outcome of reflection or as something intrinsic to our conscious mental states? We can pose this challenge as a question about whether reflective self-consciousness is the only form that self-consciousness can take. To answer this question we must ask what it is for a person to be self-conscious. This is a huge question but one way of answering it would be to determine whether self-consciousness should be understood as a relational property of a person or as an intrinsic

²⁹ For the distinction between “personal level” explanation and “sub-personal level” explanation see Dennett (1969). The distinction is discussed and put to work by Hornsby in her (1997, ch.7 & 10)

Consciousness, Self-Consciousness and First-Person Knowledge

property of a person. If the property of being self-conscious is understood as a relational property it might be thought that this property is instantiated when a certain relation holds between her mental states. Perhaps the relation in question holds when her conscious mental states are accompanied by higher-order representations of some kind which target a conscious mental state the person is in at the time. Alternatively some of her mental states might be directed upon themselves in the way that Brentano describes. We shall see in a moment that the phenomenologist argues against both of these options. In doing so he rejects the conclusion that self-consciousness is a relational property concluding instead that it must belong to certain mental states intrinsically. The soundness of the phenomenologist's reasoning here once again depends on the claim that a subject cannot become self-conscious through reflection. Let us examine then how the phenomenologist argues for this claim.

The argument I shall develop is based on a passage in the Introduction to *Being and Nothingness*, where Sartre considers whether self-consciousness might be identified with self-knowledge.³⁰ By "self-knowledge" I mean knowledge of propositions of the form "I am F", where "F" picks out some conscious state or activity of a subject and "I" refers to the subject that is thinking she instantiates F. Sartre is considering theories which take self-consciousness to be a variety of propositional knowledge, where the objects of this knowledge are propositions of the form I am in F. Self-knowledge needn't be conceived of as the outcome of a subject self-ascribing a conscious state or activity. If there are conscious states of the kind Brentano describes – states which are simultaneously directed upon themselves and at some state of affairs in the world – these states of mind will also qualify as instances of self-knowledge. These will be states of mind which are such that when they occur the subject will have knowledge of their occurrence.³¹ I shall

³⁰ Sartre (2000, xxviii-xxx)

³¹ Hossack (2002) defends a Brentanian account of consciousness according to which a mental state M is conscious if the subject has knowledge of M's occurrence. According to Hossack's theory every conscious mental state (or as Hossack puts it "every experience and every action") counts as an instance of self-knowledge: a mental state is conscious only if the

Consciousness, Self-Consciousness and First-Person Knowledge

take the target of Sartre's argument to be any account which treats self-consciousness as a relational property. Both Brentanian theories and higher-order theories of consciousness fit this description. Sartre is offering an argument for the conclusion that self-consciousness couldn't possibly come about through a subject taking up a relation to his own mental states. Hence self-consciousness shouldn't be conceived of as a relational property but should instead be taken to be an intrinsic property of our conscious mental states.

Sartre's argument bears an uncanny resemblance to an argument that can be found in Aristotle. In *De Anima* 3.2, Aristotle notes that we can perceive that we perceive, and wonders how this is so.³² He goes on to argue that either we perceive that we perceive by one and the same perceptual act or we do so by means of a distinct act of perception. Aristotle argues that if we perceive that we perceive by means of distinct act then we must posit a further act to explain how we perceive the first act. Once we have posited a further act we will need yet another perceptual act to explain how this further act is perceived. Thus we have the beginning of a regress. To end the regress we must posit a perceptual act that perceives itself. Once we concede that some act can perceive itself, we might as well say this from the outset, thus avoiding the threat of the regress.

Sartre takes a similar argument and applies it to views which seek to identify self-consciousness with self-knowledge.³³

subject knows of its instantiation.

³² *De Anima* (425b12-25). There is some disagreement among commentators about whether Aristotle thought there was a capacity responsible for a subject's perceiving that she perceives or whether Aristotle should be read as claiming that there is an activity in virtue of which we perceive that we perceive. For discussion and a defence of the latter reading see Caston (2002). I shall follow Caston in reading Aristotle as discussing the activity of perceiving because to do so fits with my current concerns, but I must confess that I don't have sufficient knowledge of what Aristotle says elsewhere to begin to defend such a reading.

³³ I don't mean to suggest that Aristotle and Sartre return the same answer to the question how is it that we can perceive that we perceive. Caston (2002) argues that Aristotle held a view of consciousness much closer to that of Brentano, according to which every perceptual act is directed upon itself. Kosman (1975) defends a reading of Aristotle that stresses the continuity between his position and that of the phenomenologists, in particular Sartre.

Consciousness, Self-Consciousness and First-Person Knowledge

“The reduction of consciousness to knowledge in fact involves our introducing into consciousness the subject-object dualism which is typical of knowledge. But if we accept the law of the knower-known dyad, then a third term will be necessary in order for the knower to become known in turn, and we will be faced with this dilemma. Either we stop at any one term of the series – the known, the knower known, the knower known by the knower, etc. In this case, the totality of the phenomenon falls into the unknown; that is we always bump against a non-self-conscious reflection and a final term. Or else we affirm the necessity of an infinite regress (idea, ideae *ideae*, etc.) which is absurd...Consciousness of self is not dual. If we wish to avoid an infinite regress, there must be an immediate, non-cognitive relation of the self to itself.” (Sartre, 2000: xxviii-xxix)

Sartre’s argument in this passage mirrors Aristotle almost exactly, but before we can see how it will be helpful to reconstruct his argument somewhat.

First off it should be noted that while Sartre frames his discussion in terms of consciousness, it is in fact self-consciousness that he is discussing. In the previous section, where I introduced the phenomenologist’s conception of consciousness, I noted that in general they (intentionally) do not distinguish consciousness from self-consciousness. There we saw Sartre claim that to be conscious of something is also to be self-conscious. Moreover, there is further evidence that Sartre is in fact talking about self-consciousness at the end of the passage where he says “consciousness of *self* is not dual”.

As I have already explained, I take the target of the argument in this passage to be theories which take self-consciousness to be a relation between a subject and her conscious mental states. When Sartre talks of “the subject-object dualism...typical of knowledge” and the “knower-known dyad”, I take him to be referring to the relation of representation a subject can stand in to her conscious mental states. This representation might be a second-order representation in virtue of which a subject represents that she is in an intentional state M, or it might be a representation of the kind posited by Brentano – a representation that is somehow directed upon itself in such a way as to represent its own occurrence.

Consciousness, Self-Consciousness and First-Person Knowledge

Sartre says that any view of this kind will fall foul of an infinite regress. He begins by noting that it will not do to conceive of the representation as a two-place relation between a subject and a mental state she is in. It is necessary he tells us, for such an account to introduce “a third term” in order for “the knower to become known”. I take him to mean that it is necessary to introduce an additional layer of representations in order for the subject to be conscious that he himself is in a mental state M. To introduce this additional layer of representation is, Sartre claims, to take the first step on a path that leads to an infinite regress. We might challenge Sartre on this point, for we have seen that a Brentanian will conceive of self-consciousness as self-knowledge, while nevertheless offering a one-level account of self-consciousness. We shall see however that Brentano’s account of self-consciousness is vulnerable to a different objection, so let us ignore this possibility for the time being.

Sartre doesn’t say why we need to introduce an additional layer of representations, but we can fill in his reasoning here without too much trouble. Sartre is rejecting a proposal that takes self-consciousness to be a two place relation of representation a subject stands in to a mental state. We can see why this cannot be right by considering how a subject P might stand in a relation of representation to another subject’s Q’s mental states. But of course standing in such a relation of representation wouldn’t make P self-conscious of being in Q’s mental state. Suppose the state in question is a pain state. It is certainly true that we can be conscious of another person’s being in pain by representing that this person is in pain. However when we represent that another person is in pain this doesn’t for the most part (ignoring the possibility of sympathetic pain) make us conscious of ourselves being in pain. The pain I thereby experience belongs to the other person and not to me. Any account which identifies self-consciousness with a two place relation between a subject and a mental state will fail to distinguish the case in which a subject is representing his *own* mental states from the case in which he is representing some other subject’s mental states. A subject will not be self-conscious by

Consciousness, Self-Consciousness and First-Person Knowledge

representing a mental state M, he must also represent that *he* is representing M.

Now recall Aristotle's argument. If we explain how we perceive that we perceive by introducing two perceptual acts, the second of which is directed at the first, we will have to introduce a third perceptual act to explain how the second perceptual act is perceived, and so on *ad infinitum*. Sartre applies this same argument to the account which equates self-consciousness with self-knowledge. I have just explained why we must introduce a further layer of representation (Sartre's "third term") if we are to account for the knowledge a subject has that he is in a mental state M. Sartre claims that once we have introduced a third layer of representations we will have to introduce a fourth layer of representations to explain how the subject knows that he knows he is in M, and so on *ad infinitum*.

The way to terminate the regress is to say that there comes a point at which the subject no longer needs to know that he knows, or in Aristotle's case, that the subject no longer needs to perceive that he perceives. Each additional layer of representation has been introduced to explain how a subject can *know* that he is in a conscious state M. If we say that there is a point at which we no longer need to introduce an additional representation Sartre thinks this amounts to conceding that there is a point at which "the totality of the phenomenon falls into the unknown". Just a few passages earlier in his Introduction Sartre has rejected the possibility that consciousness could be "ignorant of itself" as absurd. A consciousness ignorant of itself is unconscious, and Sartre (following Descartes) was of the opinion that there is nothing in the mind of which a subject is not conscious. There is much about ourselves which, according to Sartre, we *choose* not to see, but there is nothing about ourselves of which we are entirely ignorant.

Nowadays, knowing all that we do from cognitive psychology, few of us will be prepared to join Sartre in his rejection of the unconscious. For those willing to allow the existence of unconscious mental states there is then, a way out of the regress Sartre has described. Whether it robs Sartre of his argument for the dependence thesis remains to be

Consciousness, Self-Consciousness and First-Person Knowledge

seen.

Before I consider this possibility in more detail we should note that it is not a possibility that Brentanian's can take up. Brentano also employs an argument along the lines of Aristotle's in arguing against higher-order accounts of consciousness which appeal to unconscious mental states to block the regress that would otherwise ensue. Brentano thinks he has found another way of blocking the regress without introducing anything like an unconscious mental state. He appeals instead to mental states which are directed at themselves and as a result represent their own occurrence. Brentano may have found a way out of the regress without appealing to unconscious mental states but it has been argued that he does so only by presupposing what he is trying to explain.³⁴

Recall that according to Brentano every mental state is directed towards two objects, a primary object which is usually some worldly state of affairs and a secondary object which is the mental state itself. Consider Brentano's example of hearing a tone. Brentano claims that when I hear a tone, I am conscious of the tone and I am also coconscious of my experience of hearing the tone. Now we might be entitled to ask whether I am also conscious of this coconsciousness. Brentano replies that I am:

“In the same mental phenomenon in which the sound is present to our minds we simultaneously apprehend the mental phenomenon itself. What is more, we apprehend it in accordance with its dual nature insofar as it has the sound as content within it, and insofar as it has itself as content at the same time.” (Brentano, 1995: 179)

According to Brentano a single mental state can make me:

- (1) Conscious of a tone.
- (2) Coconscious of my hearing the tone.
- (3) Conscious of my being coconscious of my hearing the tone.

³⁴ See Zahavi (1999: 30-1). Zahavi credits this argument to the Heidelberg school of philosophers, specifically to Cramer and Pothast.

Consciousness, Self-Consciousness and First-Person Knowledge

Brentano cannot allow that (3) is a distinct object from (2), for if he does he will have a new regress on his hands: he will have to introduce a further conscious state which is conscious of being conscious of being coconscious. Brentano must say that when I am coconscious of my hearing the tone I am also conscious of my being coconscious. Then self-consciousness isn't something that is brought about by a mental state's being directed upon itself. The secondary object is instead already in "possession of self-awareness" as Zahavi puts it. What (3) describes is a state in which I am conscious of the tone and of myself hearing the tone. To be conscious of myself hearing the tone is to be self-conscious. Brentano must say that this self-consciousness form a part of the secondary object which a mental state is directed towards when it represents its own occurrence. Supposing this possibility is coherent, something I will not challenge here³⁵, it seems that Brentano finds a way out of Aristotle's regress only by giving us a circular explanation. I conclude that a Brentanian account of self-consciousness as self-knowledge fails. Let us turn our attention now to higher-order theories.

Aristotle and Sartre's argument will only prove effective against higher-order theories if it is assumed that all mental states are conscious. This is something that all proponents of higher-order theories deny. Thus to pursue the regress argument against these theories would require us to defend the claim that there are no unconscious mental states. I am

³⁵ Bell (1990: 19-23) argues that Brentano's position must ultimately collapse into incoherency. Brentano tells us that a representation of a sound without a representation of an act of hearing is conceivable but that a representation of an act of hearing without a representation of sound is a contradiction. Thus he conceives of a representational state's directedness upon itself as being dependent upon its directedness towards some primary object. Brentano also wants to say that it is one and the same representational state R that can be directed upon two distinct objects. The representational state R that is directed at a primary object is one and the same as the representational state R* which is directed upon itself. But on Brentano's own account of parts and wholes, if x is a part of y then x is not identical with y. It follows then that Brentano cannot claim both that R* and R are "two aspects of one and the same unitary phenomenon" the former depending on the latter, and also claim that R = R*. The claims that R* and R are two parts of the same whole contradicts the claim that the representational which is directed at a primary object is identical with the representational state directed at itself.

Consciousness, Self-Consciousness and First-Person Knowledge

convinced that such a claim must be false, so rather than attempt to show that higher-order theories fall foul of a regress I shall offer an alternative line of attack.

I will argue, in the spirit, if not the letter of phenomenology, that reflective self-consciousness has certain features which cannot be explained by any act of reflection. These are features that can only be explained by granting that conscious mental states are already self-conscious prior to any act of reflection. The features in question will already be familiar to us from our earlier discussion of reflective self-consciousness, but I shall nevertheless offer a brief reminder of them.

Recall that reflective self-consciousness attaches to I-thoughts, the judgements a subject makes in self-ascribing a thought or experience. We saw earlier that when a subject entertains an “I-thought” she cannot fail to realise that it is herself she is thinking about. Perry’s (1979) story of the supermarket shopper following the trail of sugar illustrates the point well. Imagine I am in the supermarket following a trail of sugar in search of the person with a torn bag of sugar in their basket. When I discover it is me with the torn bag sack of sugar and I think to myself I am making a mess, I know something that I do not know when I think that the person with a torn bag of sugar is making a mess. I know that I am thinking about myself. Of course it is true that I was also thinking about myself when I thought about the person with the torn bag of sugar, but at the time I had this thought I didn’t know I was thinking about myself. Whereas when I think that I am making a mess, I cannot fail to know that I am thinking about myself just by having this thought.

The thought that I am making a mess is an I-thought; it is a thought in which I self-ascribe an action of which I am now conscious, the action of making a mess. It is a feature of I-thoughts in general that a subject cannot fail to know he is thinking about himself. It is I-thoughts to which reflective self-consciousness attaches. Thus any adequate account of reflective self-consciousness must explain how a subject cannot fail to know he is thinking about himself when he is reflectively self-conscious.

Consciousness, Self-Consciousness and First-Person Knowledge

I-thoughts are also, to borrow Shoemaker's term, immune to error through misidentification (henceforth "IEM").³⁶ A subject cannot think, for instance, that she is experiencing pain and think something false because she is mistaken about who it is that feels pain.³⁷ Shoemaker argues that I-thoughts are not subject to error through misidentification because the subject doesn't identify herself at all when she self-ascribes a thought or experience. That is to say an individual doesn't need to first identify an individual S that is herself and then predicate of this individual the property of being in a particular conscious state. If she did need to identify herself when self-ascribing a conscious state, there would be a possibility of her erring about who it is that is in a conscious state when she self-ascribes a conscious state. Yet we have seen that when one entertains an I-thought, there is no such possibility of error. I-thoughts are reflectively self-conscious. Thus the second feature of I-thoughts that an account of reflective self-consciousness must accommodate is IEM.

Before I consider whether higher-order theories of consciousness can accommodate these two features I have taken to be essential to reflective self-consciousness, it is worth noting that a proponent of the dependence thesis is perfectly able to accommodate them. Consider first the claim that when I think I am F I cannot fail to know that the subject that I am thinking is F is myself. The dependence thesis says that F, a conscious state, is also pre-reflectively self-conscious. Recall that pre-reflective self-consciousness modifies the way in which a conscious state occurs for a subject. To say that a state is pre-

³⁶ Shoemaker distinguishes what he calls "absolute immunity to error through misidentification" and "circumstantial immunity to error". The statement "I am in pain" is absolutely immune to error because its immunity isn't contingent on anything else the subject believes. The statement "I am seeing a table" is circumstantially immune to error because its immunity is based on a belief of the speaker that she is currently seeing a table under normal viewing conditions.

³⁷ Not all the predicates I can apply to myself are in this way immune to error through misidentification. Only those predicates that I can know to apply to myself non-observationally just by instantiating them are in this way immune to error. The thought I am bleeding isn't immune to error through misidentification. I might be in a tangle of bodies and mistake someone else's blood for my own, to use an example of Wittgenstein's (1958: 66-7). Any thoughts I can form about my own mental and bodily states directly and immediately without recourse to observation will however be immune to error through misidentification (see Shoemaker (1968: 562)).

Consciousness, Self-Consciousness and First-Person Knowledge

reflectively self-conscious is to say that the state occurs in such a way that the subject is conscious of himself being in this state. I cannot fail to know I am thinking about myself when I think I am in F, because F occurs for me in such a way that I am conscious of myself having F. All I do when I judge that I am in conscious state F is make explicit the consciousness I already have of myself. I transform the implicit awareness I have that I myself am the subject of F into an explicit awareness of myself as the subject of F.

What about IEM? I-thoughts are IEM, we have seen, because I do not have to identify myself as the subject of F when I think I am F. According to the dependence thesis I do not need to identify myself as the subject of F because F is pre-reflectively self-conscious. I have F in such a way that I am conscious of myself having F. When I think that I am F I make explicit the consciousness I already have of myself. I do not need to identify which individual is the subject of F because when F occurs I am conscious that it is me having F. Thus we will see how the dependence thesis succeeds where reflection theories fail. It gives us a way of accommodating both features of I-thoughts. Let us consider whether higher-order theories can accommodate these two features beginning with higher-order perception theories.

5. Higher-Order Theories of Consciousness and the Dependence Thesis

Higher-order perception theories (henceforth “HOP” theories) say that a subject becomes conscious of being in a mental state M when M is scanned by an internal monitoring mechanism.³⁸ This monitoring mechanism is hypothesised to function in much the same manner as the senses which are directed outwards, except that it has as its function, the monitoring of a subject’s own mental states.

A proponent of HOP theory will say that a subject comes to think that she* is in M by means of a non-conscious mental state modelled on perception. Such a proposal immediately runs into trouble if we take HOP theorists at their word and model the

³⁸ See Lycan (1990/1997 & 1996) for recent defence of HOP theory.

Consciousness, Self-Consciousness and First-Person Knowledge

access we have to our own states of mind on perception as it is ordinarily understood. We can see this by considering just a few features of perception as it is ordinarily understood by us.³⁹

Each perceptual experience a subject undergoes will supply that subject with information about a multiplicity of objects. A subject can put this information to use in identifying the objects of his experience. There are several ways in which he can do so. He can identify or misidentify an object as being of a certain kind. He can identify an object perceived at one time with an object perceived at another time. This is something he might do by perceiving a resemblance between an object's properties at different times. Alternatively the object might be one that he is continuously observing in which case he will be in a position to perceptually track the object over time.

If we are to understand introspective knowledge as a form of perceptual knowledge, at least some of the features I have just described must also be true of introspection. Does introspection supply us with information about a multiplicity of objects? No, it provides us with information about one and only one object over time, ourselves.⁴⁰ Does introspection involve the identification of an object, the person doing the introspecting, in the various ways described above? No, the judgements we make about ourselves based on introspection are IEM. If introspection is to be understood as a mode of perception, it must be a mode of perception which doesn't require that we identify ourselves. For supposing that the relevant mode of perception does involve something akin to identification of a self we will have failed to accommodate IEM. HOP theories will

³⁹ Here I am indebted to Shoemaker. See in particular his (1994/1996, lecture 1) and his (1986).

⁴⁰ Martin (1997) argues that the awareness we have of our own bodies is a kind of perception which supplies us with information about one and only one object at any given moment. I do not wish to dispute this claim, though some have argued against the claim that bodily awareness is perceptual in nature (see for instance Gallagher (2003)). Perhaps the awareness we have of our bodies is best thought of as a kind of perception, but bodily awareness doesn't make us reflectively self-conscious. Even if some such case can be made for construing bodily awareness as perceptual in nature, it will be of no help to the HOP theorist who is attempting to give us an account of reflective self-consciousness.

Consciousness, Self-Consciousness and First-Person Knowledge

thereby fail one of the tests we have set for an adequate account of reflective self-consciousness.

In treating introspection as perception, we are looking for a mode of perception which is such that it supplies information about one and only one object, and doesn't require the identification of this object. This is a mode of perception unlike any with which we are familiar. Given these significant differences, one might wonder what work there is left for the analogy with perception to do. Perhaps there can be a mode of perception which supplies us with information about one and only one object – bodily awareness may satisfy this description. It is however hard to conceive of a mode of perception that doesn't require the subject to identify the object perceived. As soon as we say that the subject must identify herself when she introspects, we introduce the possibility of misidentification, but this brings the view of introspective knowledge as perceptual knowledge into direct conflict with the claim that I-thoughts are IEM.

Could HOP theories explain the knowledge I have that I am thinking about myself when I entertain an I-thought? I shall argue that they can do so only by presupposing what they are seeking to explain. Thus HOP theories fail both tests I have set for an adequate theory of self-consciousness.

When I know that I am thinking about myself I know that an identity holds between the person that is having this thought and the object of this thought. To see this, recall once again what happens when I discover that it is me making a mess in the supermarket. I realise that the person having the thought <I am making a mess> is identical with the person making the mess. Let us assume for the moment that HOP theories do take the mode of perception by which I have introspective access to my conscious states to involve identification of myself as the person making the mess, thereby rejecting IEM. Either I identify myself on the basis of some of my perceived properties in the same way as I might identify some other person, or else I receive information about myself which somehow allows me to identify myself demonstratively. Let us consider each of these

Consciousness, Self-Consciousness and First-Person Knowledge

possibilities in turn.

If a subject S is to identify herself by means of properties $[p_1, p_2, p_3 \dots p_n]$, she must know that she is the unique possessor of these properties. For we have seen how it is a feature of introspection that it gives us access to the conscious states of one and only one subject at any given moment. How can S know that she is the possessor of these identifying properties? If we say she knows she has properties $[p_1, p_2, p_3 \dots p_n]$ because this fact is something she perceives we must ask again how S knows that she is perceiving herself rather than some other individual when she perceives an individual that has properties $[p_1, p_2, p_3 \dots p_n]$? Perhaps she makes use of further identifying properties $[q_1, q_2, q_3 \dots q_n]$ in order to single out herself as the person with identifying properties $[p_1, p_2, p_3 \dots p_n]$. If so, we will face the same question once again: how does S know that she is the person with identifying properties $[q_1, q_2, q_3 \dots q_n]$? By now we should see that a regress has begun. It is a regress which can be ended only by allowing that S can know she is thinking about herself without any recourse to identification. What we are trying to explain is how I come by this knowledge. Thus it would seem that a proponent of a HOP theory must presuppose what he is seeking to explain or else run the risk of incurring a vicious regress.

There is at least one way out of this mire for a proponent of a HOP theory. He could deny that the subject needs to be in possession of any identification information in order to perceive that he* is in a particular conscious state. When we identify an object x demonstratively we do not need to make reference to perceived properties of x which distinguish x from other objects. When one identifies an object by means of a demonstrative expression such as “this” or “that” it is necessary that there be what Evans (1982) describes as an “informational link” connecting one to the object that one identifies. The subject gets information about the position of her body, her location in space, and of what she is doing at any given moment. Could it be that she can exploit information of this kind to identify that she is herself S when judging that she* is in a

Consciousness, Self-Consciousness and First-Person Knowledge

mental state M?

When I think about myself by means of “I” there is no possibility of my failing to refer to myself. This is one of the lessons we can take from Descartes’ cogito. The one proposition Descartes entertains which he concludes is not subject to doubt is the proposition “I think”. No demonstrative expression has this feature of guaranteed reference. It is consistent with any use of “this” or “that” that the entity we pick out by means of these expressions fails to exist, perhaps because it is imagined or hallucinated. Equally when a subject takes himself to refer to the same object at different times by means of a demonstrative it is possible for him to make a mistake. Perhaps the object he refers to at a later time is qualitatively identical but numerically distinct from the object he referred to at an earlier time.

These possibilities to one side there is an additional problem which renders demonstrative identification ill-suited to play the role of “I” in I-thoughts. It seems perfectly possible for S to single out an object by means of a demonstrative expression which happens to be S but for S to fail to know she is referring to herself.⁴¹ Yet the HOP theorist is appealing to demonstrative identification to explain how a subject knows that she is thinking about herself when she thinks an I-thought. Any attempt to model the subject’s identification on demonstrative identification must fail to capture what it has set out to explain. For recall that we are currently considering whether HOP can account for a subject’s knowledge that he is thinking about himself when he entertains an I-thought. I conclude then that HOP theory fails to give us a satisfactory account of reflective self-consciousness. Let us turn our attention now to higher-order thought theories (henceforth HOT theories) having given HOP theories a run for their money.

⁴¹ See Castañeda (1966) for arguments along these lines against analysing first-person reference as a demonstrative form of reference. The very same difficulty arises for the first proposal which claims a subject can identify herself by means of some identifying properties. Again it is possible for an individual to single out herself by means of some properties she has but fail to realise she is referring to herself. See footnote 20 for one of Castañeda’s examples.

Consciousness, Self-Consciousness and First-Person Knowledge

HOT theorists introduce a distinction between introspective and non-introspective consciousness which doesn't seem to be available to HOP theorists. According to HOT theory a mental state is non-introspectively conscious when it either is or could be accompanied by a HOT. HOT theorists disagree amongst themselves as to whether a mental state must actually be accompanied by a HOT or whether it suffices for a creature to be disposed to produce a HOT.⁴² I can't see how being disposed to produce a HOT could make it the case that an experience *actually* seems or feels a certain way to a subject.⁴³ For this reason I shall confine my discussion to higher-order theories which take a mental state to be conscious if the creature *actually* represents that it is itself in this mental state.

HOT theories claim to be able to recognise the special epistemic access a subject has to his own conscious mental states. This is something they take themselves to have explained by appealing to the non-inferential access that a subject has to his experiences when they are accompanied by a HOT.⁴⁴ Recall that a subject becomes introspectively conscious when he makes a conscious mental state the target of a further HOT. The HOT that accompanies a subject's conscious mental state gives the subject unmediated

⁴² Dennett (1978) and Carruthers (2000) defend a dispositionalist account of HOT theory, while Rosenthal (1986; 1990 & 1993) defends an actualist version of HOT theory.

⁴³ Carruthers (2000: ch.9, §3) tries to finesse this difficulty by appeal to consumer semantics (see Millikan (1984) for the account of consumer semantics Carruthers draws on). A consumer system is a system that uses a representation in the course of guiding behaviour, applying recognitional concepts, making inferences etc. Carruthers argues that it is presence to such consumer systems that confers on a representation its phenomenology. Particularly important in Carruthers' account is the presence of a representation to a theory of mind module which is capable of producing HOTs. Being presented to such a consumer system renders the creature capable of thinking that an experience seems or feels a certain way.

Carruthers' account responds to the problem I have just raised but still the suspicion remains that there is something magical at work in his account. How can the mere *disposition* to think that one is undergoing an experience generate an experience that *actually* seems or feels a certain way? If phenomenal character is conceived of as something of which a subject is actually conscious, I cannot see how the appeal to what consumer systems *could* do with a representation is ever going to succeed in accounting for phenomenal character. This is a difficulty that disappears once we think of phenomenology as something *with which* a creature is conscious of things in its environment. Then we can appeal to the presence of a representation to consumer systems to account for how a representation can be known from the inside. I will make a move along these lines in chapter 7.

⁴⁴ For an example of this kind of explanation see Rosenthal (1990/1997: 737-8).

Consciousness, Self-Consciousness and First-Person Knowledge

epistemic access to his own conscious mental states. All the subject need do in order to know his own conscious mental states in the first person is make a conscious mental state he is in the object of a further HOT. This HOT will give the subject direct and immediate epistemic access to his own conscious mental states.

How on this theory, do I know that it is me I am thinking about when I think I am in a conscious state F? When a person says 'I am in pain' we take this utterance of 'I' to refer to its speaker, the person that produced this utterance. Rosenthal (2002 & 2004) has argued that we should think of 'I' as functioning in an analogous way in a HOT. When I token a HOT the content of which is <I am in a mental state M>, I produce a thought that refers to me, for it is me that has tokened this thought. "I" works in such a way that each HOT I have represents a mental state I am in as belonging to me, the person that has produced that HOT. This is not to say that the content of a HOT should be taken to be <the producer of this thought is in a mental state M>.⁴⁵ When a subject thinks I am in M he doesn't *actually* refer to himself as the thinker of this very thought: the HOT he tokens doesn't describe him in this way. Nevertheless by tokening a HOT a subject is thereby disposed to think about himself in this way. This is because "I" works in such a way as to forge a connection between a token thought and the thinker of this thought. It is in virtue of this connection holding that the subject is thereby disposed to think of himself as the thinker of the thought in which "I" occurs.

Can such a view recognise the phenomenon of IEM? Recall that I-thoughts are IEM insofar as it is not possible for me think that I am in a conscious state F and think something false because, although someone is in F, it is not me that is in F. Rosenthal distinguishes between a strong form of IEM and a weaker variety. A proponent of strong IEM will claim that when I think I am in F, I cannot be wrong about whether it is *I myself*

⁴⁵ Rosenthal (2002) notes that if every HOT was about itself, we would have to say that "each HOT makes one conscious of that very HOT, and hence that all HOTs are conscious." (p.331) Yet Rosenthal wants to say that we are only conscious of a HOT when we make it the object of an act of reflection. This objection was first raised by Natsoulas (1993).

Consciousness, Self-Consciousness and First-Person Knowledge

that is in F. The weaker form of IEM says that when I think I am in F I cannot be mistaken that I am the individual who *thinks* he is in F. Rosenthal rejects strong IEM. He claims that if I am labouring under the misconception that I am Napoleon, I will misidentify myself when I token an I-thought. Since in these circumstances I will be mistaken about who it is that I am, Rosenthal infers I must be able to make a mistake about which person it is I am thinking about when I token an I-thought. Suppose I think I am in pain. If I think I am Napoleon, Rosenthal claims I will misidentify myself when I think I am in pain for I will think Napoleon is in pain. It is on this basis, so far as I can tell, that Rosenthal rejects the stronger form of IEM.

Rosenthal is right that we can sometimes misidentify ourselves, as I would do if I were to suddenly think that I am Napoleon. I am not persuaded however that he has established we can misidentify ourselves when we token an I-thought. In the case Rosenthal describes I misidentify myself by thinking about myself *as* Napoleon, but I do not think Napoleon is in conscious state F when I think I am in conscious state F. As has often been noted I can still use “I” to think about myself even if I suddenly find myself amnesiac, I can for instance think to myself <who am I?>. ⁴⁶ Thus even though I can be confused about my own identity it doesn’t follow that when I use “I” either in thought or talk, I will be mistaken about who it is I am referring to.

Rosenthal thinks that when I employ the first-person pronoun to refer to myself, it is not my use of “I” that singles me out from other individuals. Rosenthal claims that there is no single way in which we identify ourselves when we refer to ourselves in the first-person. Instead we appeal to a whole range of considerations, such as bodily features, events from our past, various psychological propensities and dispositions. ⁴⁷ What unifies these disparate factors is that in each case I believe something about myself in the first-person, but it is not the fact that I think about myself in the first-person in each of

⁴⁶ The example is Anscombe’s (1994).

⁴⁷ See Rosenthal (2002: §IV & 2004 §V)

Consciousness, Self-Consciousness and First-Person Knowledge

these cases that does the work in singling me out from other individuals. Rosenthal seems to think that I identify myself by means of all the propositions I believe about myself. Presumably Rosenthal will say of the amnesiac case that when I use “I” to refer to myself in such circumstances I can single out the thinker of the thought in which a token of “I” occurs but I cannot single out myself.

One will only accept Rosenthal’s argument against the stronger form of IEM if one grants to him that we do indeed identify ourselves by means of these various beliefs. For then it will be possible for us to hold false beliefs about ourselves, and hence misidentify ourselves as a consequence of these false beliefs. I can’t see why any advocate of strong IEM will go along with Rosenthal’s account of self-identification. They will deny that in order to know we are thinking about ourselves when we token an I-thought we must identify ourselves by means of the various beliefs we have about ourselves. We know we are referring to ourselves when we token an I-thought because to token an I-thought is to think about oneself self-consciously. Thus a proponent of strong IEM will reject any role for first-person beliefs of the kind Rosenthal describes. If Rosenthal is to justify his rejection of strong IEM he needs to persuade us of his own account of self-identification. All he has shown so far as I can tell is that there is an alternative position one can take on self-identification to that taken by proponents of strong IEM. I cannot find any reason for preferring his account to the one I have endorsed above.

Rosenthal does allow that I-thoughts have what he describes as a “weak” form of IEM. A subject cannot think that she is in a conscious state F but be mistaken about who it is that *thinks* she is in F. Rosenthal explains: “The error I cannot make is to think, when I have a conscious pain, for example, that the individual that has that pain is someone distinct from me...” (2004: 171). Rosenthal explains this immunity to error as being a consequence of the phenomenon of first-person reference.⁴⁸ The content of my thought that I am in a mental state M, refers to me, the individual I could describe as the

⁴⁸ See Rosenthal (2002: 344-9 & 2004: 168-176, pp.173)

Consciousness, Self-Consciousness and First-Person Knowledge

thinker of this thought. If when I think I am in pain, I am also thereby disposed to think that I am the individual who thinks he is in pain, it is not possible for me to make the mistake of thinking that the individual who thinks that he is in pain is someone else other than me. This is because having the thought I am in pain makes me disposed to think that I am the person having this very thought. Weak IEM falls out of the feature of I-thoughts whereby just by having such a thought I am disposed to think that I am the person having it.

HOT theory does seem to allow for weak IEM at least. Is accounting for weak IEM enough to explain what it is for a subject to be reflectively self-conscious? Strong IEM was introduced to explain one of the ways in which when a subject tokens an I-thought she is self-conscious. It was argued that the subject doesn't need to identify herself because when she thinks I am in M she is conscious of herself, she is self-conscious. Weak IEM falls out of the disposition a subject has when he thinks about himself in the first person to think that he is the thinker of this very thought. Thus Rosenthal will have to say that the subject is conscious of himself – he is self-conscious – in virtue of instantiating this disposition.

Rather than press Rosenthal on this claim, let us consider whether his account can handle the second feature of reflective self-consciousness I described. Recall that when I think <I am making a mess>, I know that the person that is making a mess is identical with the person having this thought. Can HOT theory account for this knowledge? I think not.

The account Rosenthal has given of self-identification is such that when I think “I am F” I am disposed to think <the thinker of this very thought is F>. I use the various beliefs I have about myself to determine that I am the thinker of this very thought. The trouble is that these various beliefs leave plenty of room for me to wonder whether it is really me that is F. This conflicts with my claim that I-thoughts do not leave room for doubt of this kind. Just by my having such a thought, I know I am identical with the

Consciousness, Self-Consciousness and First-Person Knowledge

thinker of this thought. This is knowledge that HOT theory cannot account for. The account HOT theory has given of self-reference will always leave it open for me to wonder whether it is me that is F when I think <I am F>. This is a question which I have argued shouldn't make sense given an adequate account of reflective self-consciousness.

I conclude then that HOT theory also fails at least one of the tests I have set for an adequate theory of reflective self-consciousness. We have considered three different versions of the reflection theory and found each of them wanting. I take this to supply support for the dependence thesis. I will finish up by briefly considering the relevance the argument of this chapter has for the claim that naturalists cannot admit subjective facts. I will argue that the account we have given of phenomenal character can help to make this claim more precise. First of all let me briefly summarise the argument of this chapter so far.

The aim of the chapter has been to explain why there is something rather than nothing it is like for a subject to undergo a conscious experience. I began by arguing that a state is like something for a subject – it has a “phenomenal character” – when it can be known in the first-person. A state can be known in the first-person I suggested only because it seems or feels a certain way to its subject.

We then sketched two possible accounts of phenomenal character, both consistent with an intentionalist account of phenomenal properties. The first claimed that a state's phenomenal character is something of which a subject is conscious. The second claimed that phenomenal character is something with which a subject is conscious. I went on to argue that higher-order theories subscribe to the first account of phenomenal character and intentionalists to the second. This left us needing to choose between these two distinct accounts of phenomenal character.

In the latter half of the chapter I introduced the phenomenologist's account of consciousness as pre-reflective self-consciousness. According to this account, a state has phenomenal character when a subject implicitly represents that he himself is in this state.

Consciousness, Self-Consciousness and First-Person Knowledge

I argued for the phenomenological account of phenomenal character over that of the higher-order theorist by defending the dependence thesis: the phenomenologist's claim that a state can become reflectively self-conscious only if it is already pre-reflectively self-conscious. We saw earlier in the chapter that when a subject has first-person knowledge of a conscious state or activity he is self-conscious. The self-consciousness that accompanies first-person knowledge is reflective self-consciousness. According to the phenomenologist then, a subject can have first-person knowledge of a mental state only if M is pre-reflectively self-conscious. We have also argued that a state can be known in the first-person only if it seems or feels a certain way to a subject, only if it has phenomenal character. By defending the claim that reflective self-consciousness depends on pre-reflective self-consciousness we get an argument for the phenomenologist's account of phenomenal character. By arguing that higher-order theories cannot account for reflective self-consciousness we have also in effect argued that they cannot account for first-person knowledge. I have claimed this is precisely what they must explain if they are to account for phenomenal character. Thus the argument of this section has not only supplied a defence of the dependence thesis. It has also given us a reason to prefer the phenomenologist's account of phenomenal character over its rival. Once we think of phenomenal character as that which makes it possible to know one's own mind in the first-person, it will follow that only the phenomenologist can explain what it is for a conscious state to have phenomenal character. For only the phenomenologist can explain how a state can be known in the first-person.

I will finish up by returning to the argument that naturalists cannot admit subjective facts. Recall that it was argued that naturalists give us an account of reality which abstracts away from a particular subject's point of view as much as is possible. By taking science as a guide to where there is, the phenomenologist argued that naturalism must fail to acknowledge all of the ways in which the reality we experience depends on us for its existence. Standing behind this sketch of an argument is the assumption that

Consciousness, Self-Consciousness and First-Person Knowledge

science trades exclusively in objective descriptions of reality, where a description is “objective” if it doesn’t make reference to a subject or a subject’s point of view. If science gives us objective descriptions of reality, it is inferred that science must leave out from its description of reality any properties and entities the existence of which is dependent on subjects and their points of view.

For many philosophers it has seemed that phenomenal character must be an example of a property that depends for its existence on subjects of experience. Something like this view of phenomenal character as an essentially subjective property stands behind the thought that a scientific description of consciousness must fail to account for phenomenal character. The account of phenomenal character which I have advanced in this chapter tells us why phenomenal character might be thought to be essentially subjective. It also uncovers an assumption behind the thought that any property that is essentially subjective must be left out from a scientific description of reality. Let us take each of these points in turn.

The account of phenomenal character I have proposed says that an experience has phenomenal character when it can be known in the first person, and that a conscious state can be known in the first-person when it is pre-reflectively self-conscious. To say phenomenal character is essentially subjective is to claim that necessarily, if there exists an experience E that seems or feels a certain way then there also exists some subject for whom experience E seems or feels a certain way. We can see why this might be true on the account the phenomenologist has given us. Phenomenal character is essentially subjective on this theory because whenever an experience seems or feels a certain way, a subject implicitly represents himself having this experience. It is a subject’s representing himself having an experience that makes it the case that the experience he is having seems or feels a certain way to him. We must make reference to the subject having an experience in order to account for phenomenal character because it is the subject’s consciousness of himself having this experience that makes it the case that the experience has a phenomenal

Consciousness, Self-Consciousness and First-Person Knowledge

character.

I have said that we can use this account of phenomenal character to explain why phenomenal character must be missing from an objective description of reality. I take it that Sartre put his finger on the intuition that drives this thought when he said that a person's body is 'either a thing among other things, or it is that by which things are revealed to me. But it cannot be both at the same time.' (2000: 304) Sartre is claiming here that when we are aware of ourselves as subjects we are not aware of ourselves as objects, as one "thing among other things".⁴⁹ Any objective description of reality might be able to capture what we are conscious of when we are conscious of ourselves as objects. There is nothing about this consciousness that is particular to me. However when I am conscious of myself as subject of an experience I am aware of myself *qua* subject. I am aware of the ways in which my experiences seem or feel to *me*. Sartre thinks it the latter kind of awareness that cannot be described as an awareness of a thing among other things. It is an awareness that is particular to me. It is the awareness I have of the way my experiences seem or feel to me which will be missing from any objective description of the conscious mind.

This characterisation of what it is for a property to be essentially subjective puts us in a position to state the phenomenologist's argument against naturalism more precisely. The argument proceeds as follows:

⁴⁹ This is a claim that has been argued against at length by Cassam (1997). Merleau-Ponty holds a position which differs from Sartre's in important respects. He describes self-awareness as an awareness of a subject-object. Like Sartre, and unlike Cassam, he argues that the awareness we have of ourselves as subjects is qualitatively different from the awareness we have of ourselves as objects. This is a distinction he captures by distinguishing what he calls "the phenomenal body" from the "objective body" (see for instance Merleau-Ponty's discussion of Schneider in Part 1, ch.3, pp.105). Of course Merleau-Ponty doesn't think that the phenomenal body is a distinct body from the objective body. His view is that each of us has one body which is presented to us in two different ways. What Merleau-Ponty argues against is any attempt to reduce the phenomenal body to the objective body. In other words, he argues against views which would try to identify the ways in which I represent my bodies when I am aware of my body as subject the ways in which I represent my body when I am aware of it as an object.

Consciousness, Self-Consciousness and First-Person Knowledge

- (P1) Naturalism gives us objective descriptions of reality.
- (P2) To be aware of oneself as subject is *not* to be aware of oneself as an object.
- (P3) A conscious state S has phenomenal character only if the subject that is in this state is aware of herself as being in S – if she is aware of herself *qua* subject of S.
- (P4) The objective descriptions of the natural sciences can represent the awareness we have of ourselves as an object.
- (P5) No objective description can represent the awareness we have of ourselves as subjects.
- (P6) Naturalism will leave out from its description of reality, the awareness we have of ourselves as subjects.
- (CON) Naturalism will leave phenomenal character out from its description of reality.

The next chapter will argue that naturalism can admit subjective facts. It will argue against the claim that there is something particular to me about the experiences I undergo that cannot be given an objective description. Thus I will be rejecting (P5). Once we have shown that naturalism can admit subjective facts, this will clear the way for me to make use of the phenomenologist's descriptions of consciousness in developing a naturalistic account of phenomenal consciousness.