

## SELF-ASCRPTIONS OF BELIEFS

Pascal Engel

Université de Caen, CREA, and Institut universitaire de France  
*Prepared for a lecture in Bologna never delivered, unpublished 1997*

“Cartesianism” in the philosophy of mind is most often understood as the view that it is of the essence of the mind that each mind has a special, privileged access to its own contents. It is often called the view according to which there is a “first-person authority” with respect to the contents of our own mental states, an authority that we do not have in the second or in the third person, when we ascribe mental states to others. Cartesianism is also associated to another thesis, that first-person authority is “infallible”, or “incorrigible”, in the sense that we cannot be mistaken on the contents of our own minds.

So characterised, Cartesianism has had a rather bad reputation in contemporary philosophy, at least since Wittgenstein and Ryle. Ryle, for instance, argues in *The Concept of Mind*, that there is no special first-person authority, and that our access to the contents of our own minds has no special privilege over our access to the contents of the minds of others, and that it is no less fallible. This claim has recently been revived by Daniel Dennett (1991) in his attacks against what he calls the Cartesian theater of consciousness. Other philosophers have also suggested that externalism in the philosophy of mind, the view according to which the contents of our thoughts are at least partly determined by external features of a thinker’s environment, undermines first-person authority. Finally, Cartesianism is associated with mind-body dualism, and it has been attacked by philosophers of materialist or physicalistic persuasion. All these anti-Cartesian arguments are supposed to lead to the thesis that it is in no way the essence of the mind that it can be accessed from a first-person point of view.

But the fact that we have a privileged access to our own minds seems to die hard. It does seem to us that we can have a special kind of self-knowledge of our own mental contents, and that we do not have such privileged knowledge with respect to the minds of others. The view that we could have no privileged first-person authority also seems to fly in the face of some obvious facts. For instance when we communicate our beliefs to others, we cannot say that we believe something with the intention of conveying this belief if we do not believe that we believe it. As G.E. Moore pointed out long ago, it would be odd to say “P”, that is to express linguistically that one believes that P, while at the same time saying that one does not believe that P, hence that one does not believe that one believes that P. This why saying, for instance, “The train for Bologna starts

at 8.p.m, but I do not believe that it does” seems to be self-defeating, and also to defeat the purported act of communication. It also seems that when we deliberate about what to think or to do, we need to know what we believe, in order, in particular, to be able to revise our own beliefs in presence of contrary evidence. A rational individual should know about his own beliefs, and the anti-Cartesian view that he cannot know that, or that he is in no way privileged in this, seems to undermine the very possibility of being rational in this ordinary sense.

I think that such facts render implausible the anti-Cartesian view. But it does not follow that Cartesianism, as I just characterised it, is justified. It could well be that the phenomenon of first-person authority is compatible with our not being infallible about the contents of our own minds. In fact Cartesianism is not a unified view. We should in particular distinguish the following claims, although they are often conflated:

- a) that the mind has a privileged access to its own contents
- b) that this access yields infallible knowledge
- c) that it is of the essence of the mental that it can be accessed, in a privileged way, in the first-person
- d) that there are two distinct substances, mind and body.

In what follows, I propose to argue that a) and c) are true, although b) is false. I also believe that d), mind-body dualism, is false, and that its falsity is compatible with the truth of a) and c), but I shall not deal with this. I shall try to defend, with a) and c), a moderate form of Cartesianism, by arguing that self-knowledge is possible. The difficult question is: *why* is it possible? Barely formulated, my answer is very simple, even simple-minded: it is possible because we have the capacity of entertaining second-order beliefs, beliefs about our own beliefs, and because this capacity is, in a sense to be specified, self-warranted. Again, the difficult question is: what is the source of this warrant?

## I

If we do not simply deny that self-knowledge is possible and exists, we must give an explanation of this fact. But the philosophers who want to make this fact compatible with a broadly anti-Cartesian view of the mental, i.e who want to reject theses a) through d), must account for the appearance of self-knowledge, while at the same time denying that it is infallible knowledge. An answer which is tempting, along this line, is that self-knowledge is simply based upon a capacity of *introspection*, understood as the capacity to observe the contents of our own minds. Introspection, on this view, is a special capacity that we have to discover what happens in us when we have certain mental states, such as pains,

sensations, beliefs, desires, and other propositional attitudes. If we take the notion of an inner sense at face value, introspection is simply a certain form of perception. To hold this view, one need not, however, endorse the thesis that there is a special organ of inner sense, which would, as it were “scan” the contents of our minds. Thus David Armstrong, one leading proponent of this view, says:

“Eccentric cases apart, perception, considered as a mental event, is the acquiring of information, or misinformation about our environment. It is not an “acquaintance” with objects, or a “searchlight” that makes contact with them, but it is simply the getting of beliefs. exactly the same must be said of introspection. It is the getting of information or misinformation about the current state of our mind.” (Armstrong 1968, p.326)

On this view, self-acquaintance or introspection is analogous to sense-perception, because just as sense-perception allows us to acquire beliefs about our external environment, introspection allows us to acquire beliefs about, so to say, our internal environment. The important fact, for those who hold this view, is that one does not need to have beliefs about our own beliefs, i.e to believe that we believe something, in order for the introspective mechanism to yield beliefs about ourselves. Armstrong also emphasizes the fact that the beliefs that we thus get about our mental happenings need not be about a special object, a *self*. So for him introspection is perfectly compatible with our being acquainted with a bundle of mental items, composing a Humean or scattered self. This is why he holds that the capacity of introspection is compatible with the denial of the existence of a mental substance or thing which would be the object of these perceptions. But I said that I would not deal with this metaphysical aspect of the problem.<sup>1</sup>

The important point, again, on this view, is that there is a belief-producing mechanism, which produces beliefs “about oneself”. Saying that it is a mechanism implies that it is a causal mechanism, hence a contingent one. The fact that this mechanism causes some beliefs about the existence of some state of affairs, the mental items which populate our minds, means that the cause of these beliefs, and their effects, the beliefs themselves, are logically independent from each other. In other terms, just as sense-perception yields beliefs about external states of affairs though a causal regularity, inner-perception yields beliefs about internal states through a causal regularity. (see also Mellor 1991)

We are now in position to understand why, on this perceptual model of self-knowledge, self-knowledge seems to be warranted. It is because the mechanism operates, in general, in a reliable way. “In general”, because it is not without exceptions. In some circumstances, the mechanism can fail to produce true informations about oneself; it can, as Armstrong says,

---

<sup>1</sup> On this point, see Engel, to appear

produce also *misinformation*, just as perception can lead us to mistaken beliefs. The causal mechanism of introspection is reliable, but it is not infallible. Hence the Cartesian thesis b) is rejected.

But we should note that the perceptual model not only rules out the infallibility thesis ,(b), but also the privileged access thesis, (a), and the thesis of the essential first-personal character of the mental, (c), as well. Why ?

Think of ordinary perception. I can, through ordinary perception, acquire the belief that this is a cat. But I can be mistaken, and wrongly judge what I see to be cat, while it is in fact a Pekinese dog. Somebody can correct my mistake, or I can correct my mistake, by attending to my first belief and revise it in the light of contrary evidence. In order to see that my perceptual belief that this is a cat is false, I must attend to my belief either in a third person way (when somebody points out to me that my belief is false), or in what we may call a derived third-person way, by ascribing to my former self, or a later stage of my mental life, a belief which my later self, or later stage of my mental life, judges false. Hence access to the truth of my perception cannot be had in the first-person way, for if I only attend to the contents of my own thoughts, I cannot judge whether they are true or false. I cannot derive anything about their semantic properties.

Now if self-knowledge is a state of the introspective or perceptual sort, it should follow that I can entertain certain beliefs, and have a certain conception of the kind of states they are, without knowing that I have them, that is without being able to self-ascribe them to myself truly. I could have these beliefs without being aware that I have them. In other words, I could say that I believe that P, without being able to say whether I believe that I believe that P.

The perceptual model of self-knowledge has, therefore, the following consequence: a person could have an acquaintance with her own states, through a causal mechanism which is generally reliable, and therefore which yields knowledge in so far as reliability can be a necessary condition for knowledge, without conceiving of herself as having these states, that is without conceiving herself as the very subject of these states. Let us call, with Shoemaker (1996) a creature would would instantiate this possibility a creature which is “self-blind”. Can there be such creatures and could we be creatures of this kind?

## II

I have already suggested above that a rational being, a being who would be able to revise its own beliefs, to deliberate about them or about a course of action, and to communicate his own beliefs to others, needs not only to have first-order beliefs and desires, but also to have second-order

beliefs and second-order desires, what we may call reflexive beliefs or goals. But these are only intuitive claims. In order to establish them, we need something like an argument to the effect that a creature who would be self-blind could not appropriately qualify as a rational creature. I shall now state such an argument, which I essentially borrow from Shoemaker (1996). We may call it the argument from Moore's paradox.

"Moore's paradox", consists in remarking the paradoxical, or at least the anomalous character, of sentences of the form: "P, but I don't believe that P". The reason why it is not a paradox in the usual sense, is that the first conjunct does not formally contradict the second, since both conjuncts might be true. It may well be true, for instance that the earth is round, but that I do not believe it. But the sentence is nevertheless contradictory, since the first conjunct implies that I believe that it is true, whereas the second denies this. As Wittgenstein (1980) said in commenting upon Moore's paradox, the paradox shows something about the "logic of assertion": asserting that P is the usual way of expressing that one has the belief that P, and therefore denying that one has the corresponding belief seems to contradict the belief expressed by the first conjunct. But, according to Wittgenstein, it is not a *logical* contradiction, because such sentences as "I believe that P" are not descriptions of one's state of belief, but expressions (*Ausserungen*) of them. When ascriptions of beliefs are made in the second or in third person, there is no corresponding oddity. For instance there is nothing paradoxical in saying "The earth is round, but he does not believe it", because the "logic", or the "grammar" of such third person ascriptions of beliefs is such that they are actually *descriptions*, and not *expressions* of beliefs. The paradox is not a semantic one, but a pragmatic one. We may put a similar point along Gricean lines. We could say that someone would asserts that P intends to convey to an audience that he believes that P, and has the higher-order intention of intending the audience to recognise his first-order intention. Assertion is an action done with the intention of producing the belief that one has the belief. The moorean sentence defeats this purpose, and therefore does not successfully convey the intention conveyed by the first conjunct; indeed it cancels it.

All this is common wisdom about Moore's paradox. What is less often noted (but see Heal 1994), is that Moore's paradox is not only present at the level of language, or of the linguistic expression of thought, but at the level of thought or belief itself. As I remarked above, if we take the first sentence of the Moorean conjunction "P, but I do not believe that P" to express the belief that P, and the second sentence to express disbelief that P, there is no contradiction. There is no contradiction, because I may well believe something, and disbelieve it. For instance, I may at one time believe that Chirac is a good president, and fail to believe that at a later time, or I may discover that I have both beliefs, without having noticed it until now. But of course there is a contradiction if I have both beliefs, and

if I am aware that I have them, and if I go on believing them while being aware of this. In other terms, if the Moorean sentence is understood in this sense:

“ P and I don’t believe that P, and I *believe* that I believe that P and that  
I don’t believe that P”

then there is a genuine contraction. In other terms, the subject who entertains the beliefs expressed by the Moorean sentences can entertain such beliefs, but he cannot believe that he has these beliefs, unless he explicitly contradicts himself. In this sense Moore’s paradox is a paradox because there cannot be such beliefs as those expressed by the sentences, not because there is something wrong in their linguistic expression.

But we have not yet explained why there is a constitutive connexion between believing and believing that one believes. Nor have we explained what it is to believe that one believes. There are two possible explanations.

One explanation involves the notion of conscious belief. We can say that asserting a sentence implies that one is conscious of of the belief that it expresses. On one analysis of conscious belief, a conscious belief is simply a second-order belief, a belief that one has the first-order belief (Rosenthal 1978). The explanation of Moore’s paradox would then be that the content of the paradoxical sentence cannot be *consciously* believed. We can conceive of conscious belief as a sort of mental counterpart of assertion, a mental assent to a given content which is presented, in some way, to our mind. Thus the explanation of the constitutive link between assenting that P and assenting to “I believe that P would be that if one assent to the first content, one assent to the second.

But this first explanation ignores the fact that we can believe that P, but nevertheless do not assent to our believing that P. Ordinary cases of self deception provide numerous examples. I can believe that I am not as good a philosopher as Kripke, but nevertheless not assent to this. This explanation leaves out the fact that much of our beliefs are not conscious. But here we should be cautious. If by “unconscious belief” we mean beliefs such as Freudian beliefs, which could be repressed in some obscure part of our minds, then there does seem to be possible to have such thoughts as those expressed by sentences such as “P but I do not believe that P”. The first belief could be stored in one part of our mind to which we have conscious access, whereas the second could be stored in another part of our mind to which we do not have conscious access. Thus a psychoanalyst, for instance, could lead me to express such beliefs as “ I want to kill my father, but I do not believe that I want to kill him”. In such cases I would really be self-blind, in the sense above, that is in the sense in which I could have first-order thoughts, but could not be ware of my

having such thoughts, except in a third person way. But such a possibility would beg the question, for such cases are ordinarily described as cases of irrationality, whereas what we want to say is that Moorean beliefs are not possible for a *rational person*.

Beliefs can be non conscious, while still being to some extent accessible or available. Such beliefs are usually called *tacit* beliefs. For instance I may have the tacit belief that my watch is on my wrist, without assenting consciously to it. This belief can nevertheless show itself in my actions, for instance when I look at my wrist when somebody asks me what time it is. I shall not elaborate here on the notion of tacit belief, but I shall suppose that there are such beliefs. If we just grant this fact, that assent to a belief content need not be conscious and explicit, then the claim that assenting to P implies that one believes that P, and that one believes that one believes that P, comes down to the claim that *if a belief P is available for assent*, then the belief that one believes that P is available as well (Shoemaker 1996,p.79-81).

The contrary claim just amounts to the possibility of self-blindness, the possibility of a being who would be able to have beliefs, but who would not believe, in a first-person way, that he has these beliefs, although he could believe that he has these beliefs in a third person way, for instance by gaining information about his behaviour. And, you will notice, this also amount to the possibility of a creature who could assert such sentences as “P, but I do not believe that P”, and who would not find any any impropriety in the holding of the beliefs expressed but such sentences. For instance there could be a creature who believes that P, expresses this belief by asserting that P, but discovers, in a third person way, that in fact he does not have this belief. I want to claim, with Shoemaker, that this is not possible.

Could the reason why it is not possible be that such a creature could not *utter* Moorean *sentences* ? This could not be the reason, because there is nothing which prevents such a creature from having acquired the appropriate linguistic rule “Do not utter sentences of the form “P but I don’t believe that P”. One individual could notice that, in his community, such sentences do not elicit successful communication. He could follow this rule, just in the sense of instantiating a certain regularity. He could notice, for instance, that assertion is the normal mean of expressing one’s beliefs, and assert sentences with the intention of conveying to his audience that he believes that P. but given that, by hypothesis, he is self-blind, he would lack evidence for believing that he believes that P. But if this is so, the self-blind man would be unable to use the proposition that P in his reasonings. Here is why. Normally a rational person who believes that P should be disposed to use this proposition as a premise in her reasonings, and should know that, if the proposition is true, it is of her interest to act on the assumption that it is true. And — this the important

point— such a rational person should know that to act on the assumption that a proposition is true is to act as if one believes that proposition. She should also know that it is of her interest to manifest her beliefs through assertions, if she want to communicate successfully with others. In other terms, even if we use a minimal definition of belief as a disposition to act, and if we assume a minimal notion of rationality as satisfaction of one's interests, a rational believer is a person who would act as if she believes that P. But could it be that although she acts *as if* she believes that P, she actually has grounds *not* to believe that P, and hence frame Moorean thoughts? I want to say that such a person would, at least, find an inconsistency in her own actions, not simply in her own thoughts. She would be unable to plan her own actions in the future, and to ascribe these actions to herself. In other terms, she would lack a capacity for normal rational action, because she could not find continuity in her own actions, indeed not even to find that these action are *her* actions. And if belief is a disposition to act, then she would be self-blind about her beliefs. But I have just claimed that this seems to be utterly implausible, given only minimal requirements on rationality. To the extent that a subject is rational, and possesses the concept of belief, believeing that P brings with it the cognitive disposition to believe that P, either explicitly or tacitly. In other terms, a self blind creature could not have genuinely the concept of belief.

Why does this argument cast doubt on the perceptual model of self-knowledge? Because this model implies that a being could lack the perceptual capacity of introspecting herself in order to see what beliefs he has, without being in any way impaired cognitively. But if the preceding argument is correct, this is impossible.

### III

This was a first objection to the perceptual model of self-knowledge. But we can give also another objection to it, which I shall derive from Burge's (1988). The perceptual model, as its name indicates, rests upon the assumption that there is nothing more to self-knowledge than an ability to form first order beliefs about oneself comparable to the beliefs that we can get through perception. Now if it is like perception, it follows that it can be subject to brute errors just as perception is subject to such errors. Now a most common idea about perception is that the content of our perceptual states depend essentially upon the nature of our perceptual environment, and that, if this environment would change, our perceptual states would not have the same contents, although we might be unaware of this. In particular a creature could be transported in an environment which would

look the same to her, but which would actually have different non phenomenal properties. Such cases are familiar in contemporary philosophy of mind since Putnam's Twin Earth thought experiments. I shall assume familiarity with such experiments, and simply grant the intuition which they elicit: that if a thinker had some beliefs in a particular situation, and if he were switched to an environment which would be phenomenally similar but actually different, then the contents of his thoughts or beliefs would actually be different. For instance if in a given environment a thinker's thoughts had been thoughts about twater (a substance which is not H<sub>2</sub>O but XYZ), these thoughts would be different from ordinary thoughts about water. But if a thinker believes that he believes that *this is water* in the counterfactual environment, then it seems that he should be mistaken in this belief if transported into this environment. In other terms, the "external character" of our thoughts, the fact that we could be mistaken about the contents of our thoughts because they are determined by features of our environment, seems to threaten our first-person authority on our thoughts. But, as Burge (1988) and Davidson (1987) have argued, there is no such threat. For the fact that I can be mistaken in my first order thought that this is water, while it is actually water, does not in any way impair my thinking, or my believing *that I think that this is water*. And if my belief is the belief that *this is twater*, then my second-order belief in the second, counterfactual environment will be the second-order belief that I believe that this is twater. In other terms, variations in first order contents carry over into variations of second-order contents. This amounts to saying that although I can be mistaken in my perceptual belief that this is water, I cannot be mistaken in my second-order belief that I believe this. So self-knowledge of one's own beliefs is immune to changes in environmental context, or is, as Burge says, "environmentally neutral": I may not detect the changes in my environment, but even if I don't I can't fail to think that I have the corresponding thoughts. Such second-order thoughts are, as Burge says, "contextually self-verifying".

I am, as it seems, always *entitled* to have such thoughts or beliefs about my own thoughts and beliefs.

The question is: what is the source of this entitlement? Burge's answer, as he puts it is the following:

" [our entitlement to self-knowledge] derives not from the reliability of some causal-perceptual relation between cognition and its object. It has two other sources. One is the role of the relevant judgments in critical reasoning. The other is a constitutive relation between the judgments and their subject-matter, or between the judgments about one's own thoughts and the judgments being true. Understanding and making such judgments is constitutively associated both with being reasonable and with getting them right.... To be capable of critical reasoning, and to be subject to rational norms necessarily associated with such reasoning, some mental acts must be *knowledgeably*

reviewable. The specific character of this knowledgeable reviewability requires that it be associated with an epistemic entitlement that is distinctive... there must be a non contingent, rational relation, of a sort to be explained, between relevant first-person judgments and their subject matter or truth.” (Burge 1996, p.98)

Burge explicitly rejects here what we have called the perceptual model of self-knowledge. His claim is that the reason why we are entitled to have beliefs about our own beliefs is that a being who would not have the possibility of framing such second-order beliefs would not be able to engage in critical reasoning. Critical reasoning, he says,

“is reasoning that involves an ability to recognize an effectively employ reasonable criticism or support for reasons and reasoning. It is reasoning guided by appreciation, use, and assessment of reasons and reasoning as such. As a critical reasoner, one not only reasons. *One recognizes reasons as reasons* [my emphasis]. One evaluates, check, weights, criticizes, supplements one’s reasons and reasoning. Clearly, this requires a second-order ability to think about thought contents of propositions, and rational relations among them.” (*ibidem*)

The carrying out of a proof, for instance, presupposes an ability to reasoning of this kind. A non-critical reasoner, Burge says, would reason blind, without appreciating reasons as reasons (p.99). It follows in particular that a critical reasoner needs to have the concepts of the attitudes that he has, and needs to *commit* himself towards the contents of his attitudes. If we could not be critical reasoners in this sense, “there could be no norms of reason governing how one ought check, weigh, overturn, confirm reasons or reasoning”. And there could be no such thing as epistemic responsibility, whereby we could be able to review, reflexively our reasons.

We need not enter into the details of how Burge characterises what he calls critical reasoning to understand his claim. according to him, our capacity to have second-order thoughts, reflexive beliefs, is *required* by the very possibility of engaging in such reasoning and is, therefore, its main source. Indeed the requirement goes the other way too: critical reasoning requires the capacity to frame second-order thoughts. There is a constitutive, intrinsic relation between the two.

Burge’s analysis of the source of our entitlement to self-knowledge is thus very similar to Shoemaker’s analysis. On both views, the perceptual model of self-knowledge is rejected, because it makes the source of our entitlement a merely causal and contingent source. but the source is not contingent or causal: it is indeed a necessary or a logical one. And in this sense it is *a priori*. It is an *a priori* requirement for self-knowledge that we can be critical reasoners, who are able to follow rational norms. This is why we may call this the *necessary entitlement thesis*.

## IV

I have presented two objections against the perceptual model of self-knowledge, one from Moore's paradox, the other from the compatibility of such knowledge with externalist assumptions. These objections seem to vindicate strongly the idea that the capacity to have second-order reflexive beliefs is essential for the very having of beliefs, and therefore in this respect to vindicate the Cartesian theses a), first-person authority, and c) the essential character of this authority for mental concepts (although we must note that it has only been established for beliefs and propositional attitudes, not about other kinds of mental states, such as sensations). We should also note that this is compatible with rejection of the Cartesian thesis b), for in no way Burge's nor Shoemaker's views imply that the mind is completely transparent to itself in self-knowledge, and infallible in the contents that it entertains at the first-order level. But at least there is some sort of infallibility, or immunity to error, at the second-order level.

Now, in spite of the fact that the arguments that I have presented against the perceptual model of self-knowledge and in favor of the necessary entitlement thesis seem to me convincing, they are not completely unproblematic. We should raise at least three questions for the necessary entitlement thesis.

1) One condition, as we have seen, for our capacity to self-ascribe beliefs to ourselves is that such beliefs must be available or accessible, either through some conscious assent to their contents, or because they are tacit, and, in the sense that I have given to "tacitness" here, at least subject to assent *in principle*. But what are the conditions of such an accessibility of availability? The necessary entitlement thesis is silent upon this, because it seems that asking this question would imply the giving of some causal account of the availability of contents for self-ascription, and the thesis, stating a priori constraints, does not say anything about how this availability arises.

2) This immediately raises a second question: if the conditions upon which we can assent to belief-contents matter for an account of self-knowledge, and if these conditions have a causal character, how can the necessary entitlement thesis avoid the introduction of such causal elements? And if such elements are present, does this not justify partly one of the suppositions of the perceptual model, namely that the connection between the belief-forming mechanism and the beliefs that it produces is a contingent, not a necessary or logical one?

2) This question in turn suggests a third one: given that, on the necessary entitlement thesis, a subject needs to have beliefs about its own beliefs in a reflexive way, that is to have or to possess the concept of

belief, what it it to possess such a concept? In particular is it really necessary to possess such a concept in order to engage in the activity of reasoning?

It seems that asking these questions leads us to question the correctness of the necessary entitlement thesis, and to bring us back to the possibility of self-blind creatures that we have already discarded. But it need not be so. It may well be that a correct account of self-knowledge should combine the existence of causal elements and of logical ones. This what I want to suggest. Let us postpone, for the moment, an answer to the first question, about the availability of beliefs, and let us consider the two last questions.

As we have seen from our discussion of Burge, a number of self ascriptions are “self-verifying”. For instance, to take one example from Burge, if I make such self-ascriptions as: “I judge, herewith, that there are physical objects”, such ascriptions are necessarily self-verifying, that is one cannot doubt their truth. But there are cases of self-ascriptions which are not contextually self-verifying. One is the case of ascriptions made from the existence of a memory. For instance if I ask: “What is the city to which both Garibaldi and Mussolini marched to?” and if my memory presents to me the answer: “Rome”, my self-ascription of the belief: “I believe that Garibaldi and Mussolini both marched to Rome” depends upon two things: the fact that I seem to remember that it is Rome to which they marched, and the fact that I take my memory as correct. But this is not contextually self-verifying. For my memory may be wrong. I have given this example on the hypothesis that the self-ascriber is not Italian, but, say, French. It would be different, I presume, with an Italian schoolboy, who knows, so to say, automatically, without attending to his memory, that it is indeed Rome to which Garibaldi and Mussolini marched to. In such cases, there is no need of a conscious memory nor of attending to it, just as when one is asked his phone number or his name. But here too I can misrepresent the information. there are other cases as well, which I shall not mention (see Peacocke 1996, p.121-122).

Now if we think of such cases, there is no reason to deny that the self- ascriptions are made in virtue of a causal element: it is *because* memory serves up the information that Garibaldi and Mussolini marched to Rome that he can self-ascribe this belief to himself. And it is also *because* there is some automatic access to the information in question in the Italian schoolboy that he can give the same answer. If the memory were not present, and if he were not willing to take his memory as correct, the thinker could not be entitled to the self-ascription of the belief in the first place. Therefore it seems to difficult to deny that this causal route to the availability of the belief is at least an important component of the entitlement. It follows that the necessary entitlement thesis cannot be

simply an *a priori* claim about the constitutive relationship between believing and believing that one does believe, and must include this causal condition as well. Or perhaps we should formulate the necessary entitlement thesis as a thesis to the effect that a thinker who has *normal* access to his beliefs should also be able to self-ascribe them to himself. But this normality condition just is the causal condition that we have mentioned. Nevertheless, it does not follow that we are led back to the perceptual model of self-ascription, and that we should conclude that there is nothing necessary nor logical in the constitutive link between belief and belief about belief. For it is still the case that once the thinker is caused to assent to P by his memory, and is willing to take his memory as correct, he is warranted in his ascription. Why? Because the content of the second-order belief "I believe that P" must be the same as the content of the first-order belief "P" that was entertained by the thinker. We can reproduce here Burge's argument: if the first-order belief is mistaken, then the second-order one is mistaken as well. It is the identity of the first level contents, and of the first-level concepts contained in those contents, which ensures the security of the second-order beliefs, to the effect that one cannot fail to have them, once one has them. And this is quite different from ordinary perception, because ordinary perception can lead to error. A perceptual experience is never sufficient for the correctness of a perceptual belief, whereas a second-order belief is always sufficient for the correctness of the self-ascription of it. So the recognition of the fact that there is a causal factor in our entitlement does not imply that we come back to the perceptual model. In fact the presence of this causal factor is compatible with the necessary entitlement thesis.

This answers our second question above. What about the third question raised about the necessary entitlement thesis, whether it is necessary that we have the concept of belief in order to have genuine beliefs and in order to be able to be genuine reasoners? Burge claims that it is necessary, because otherwise we would not be able to follow any norms of reasoning, and to recognise them as such. Shoemaker is less committed to the idea of there being such norms, but, as we saw, he claims that a creature who would not have such capacities would not qualify as a rational creature in her actions, in at least a minimal sense.

But this seems too strong, or too idealised. We can grant the view that to be able to reason is to be able to assess certain rational relations among one's beliefs, that is to revise them in the light of new evidence, and to act accordingly. But does it imply that the thinker has or possesses the relevant concepts of belief, or desire, or of other propositional attitudes, as the necessary entitlement thesis seems to imply? No, for there can be more primitive forms of reasoning which do not involve the possession of such concepts. Here is an example, which I adapt from Peacocke (1996, p.129), in a way to make it more politically correct than his. Your wife comes

home, and discovers that there is nothing in the fridge and that the dinner is not ready (you are a good husband, who prepares the dinner when your wife is outside). She infers, with some irritation, that you are not back home yet. But later she suddenly realises that your couple has been invited for dinner to a friend's house, and that, on this circumstance, there was no need to refill the fridge or to prepare a dinner. So your wife suspends her belief that you are not home yet, and she may later reach the belief that you are home after all, because you had no special appointment or other reason for not being there. We have here a piece of reasoning, which, we can assume, goes through your wife's mind fairly quickly, which involves beliefs and their revision, but which at no point seems to involve the need of beliefs about one's beliefs or the need to conceptualize one's having of these beliefs. Nevertheless it involves relationships of support of some beliefs from others, and the ability to revise these. This is not critical reasoning in Burge's sense, although it obviously involves some normative relationships of assessment of thought contents. The important point, however, is that it is not *blind* reasoning, as if the wife's conclusion were reached through some sort of simply causal mechanism, for the comparisons of beliefs and their revision which happens at a level situated below the level of conscious assessment of beliefs occurs also at a level which is *above* that of a purely causal mechanism involving no reflexion upon beliefs at all. The point is one need not have *fully* the concept of belief in order to engage into such reasonings. There may be second-order beliefs there, without there being any possession of the concept of belief. And I do not see any reason why creatures more primitive than adults, say infants and some animals, are incapable of engaging into such kinds of reasonings.

All this is by no means incompatible with the necessary entitlement thesis. We can say that the source of our entitlement is the result of the combination of the capacity for such elementary reasonings together with the capacity, in Burge's sense, to be a critical reasoner. The answer to our third question, therefore, could be given along these lines: a belief content can be available when at least trains of primitive reasoning of the kind suggested can occur. This does not exclude the rational or normative requirements upon belief adduced by Shoemaker and Burge, for the beliefs may be tacit. But if they are tacit in our sense, they must be at least accessible through some causal route. And this why the causal or psychological element matters. The rest of the story can be filled by psychologists, who investigate the ways through which contents can be accessed and available. It's an empirical matter what are these forms of availability, but not an unimportant one.

This, it seems to me, corrects the picture given by the necessary entitlement thesis. But it still does not explain why we are entitled to our self-ascriptions of beliefs. When I say that it does not explain it, what I

mean is this: although we have claimed that self-ascriptions of beliefs are warranted, that they should be warranted if rational thought and reasoning is to be possible, and that a capacity for having second-order thoughts is presupposed in the exercise of this reasoning, we still do not see what is the source of this warrant.

Peacocke (1996), in a paper which replies to Burge, and which has inspired much of my argument so far, gives us an answer. He draws upon, once again, Burge's argument about the environmental neutrality of self-knowledge, and upon the idea, already mentioned before that in a self ascription, the thinker must entertain, in his second-order belief, the *very same* concepts has those that he entertains in his first-order belief. He claims that, for such ascriptions to be possible, the following principle must hold:

*Redeployment Claim:* :

"The concepts (senses, modes of presentation) that feature in first-level thoughts not involving propositional attitudes are the very same concepts which feature in thoughts about the intentional contents of someone's propositional attitudes." (Peacocke 1996, p.131).

This is called the Redeployment Claim because, according to it, a thinker who ascribes beliefs to himself must redeploy, in his second-order beliefs, the very same beliefs contents as those that he deploys at the first level of his beliefs. As presented here, the Claim is about concepts and contents of propositional attitudes. But there is a semantic motivation for the claim, which is familiar from the literature on propositional attitude ascriptions. Famously Frege claimed that, in propositional attitude linguistic contexts, the expressions which figure embedded in these contexts must have their ordinary sense, although they do not have their ordinary reference. But, as many critics have noticed this raises the threat of a hierarchy of senses, for in doubly embedded contexts, such as "X believes that Y believes that *a is F*", the senses of "a" and F" in this context should not be ordinary, but other senses, and so on with multiple embeddings. A number of writers have resisted this consequence, and tried to block the hierarchy. Their motivation was the intuitive appeal of a principle that Davidson (1969) has called the principle of "semantic innocence": the meaning of words in propositional attitude contexts must be the same as those outside such contexts. In other words, it seems incredible to say that if one ascribes to Galileo the belief that the earth moves, the words "earth" or "moves" do not have their normal, customary meanings.

But apart from this semantic motivation, with which I shall not deal here, the main motivation for the Redeployment Claim is that in order for

self-ascriptions to be possible, the very same concepts as those of first order thoughts must be redeployed at the level of second-order thoughts. Peacocke gives us examples dealing with demonstratives, such as:

(1) I believe that that man over there is French

Suppose also that I believe, on the basis of evidence that I have, that:

(2) That man over there does not like croissants.

given that most French like croissants, it seems that there is a sort of inconsistency in my beliefs. But if the demonstrative “that man” does not have the same sense in (2) as in (1), the inconsistency would not go through. Indeed the inconsistency is very similar to the one that we have discussed earlier with Moore’s paradox, for we could conjoin (1) and (2) in

(3) That man over there is French, and I do not believe that he is French (for he does not like croissants)

But in order to see the inconsistency, the demonstrative concepts must be the same.

On the face of it, Peacocke’s Redeployment claim seems very plausible, and reinforces the conclusions that we had reached earlier: it is indeed a requirement that when I ascribe beliefs to myself, I must employ the same concepts as those that I employed when I entertained these beliefs, so to say, unreflexively.

One can understand that the Redeployment Claim is a requirement for genuine self-knowledge, and its insensitivity to changes in environmental circumstances. But I do not understand why Peacocke claims that this principle is “explanatory” (p.118) and “contributes to an explanation of the near infallibility of a thinker’s knowledge of the contents of his conscious beliefs” (p.147). For it is one thing to say that such a principle *has* to hold if our self-ascriptions of beliefs are to be warranted. And it is another thing to say that because it has to hold, our self-ascriptions *are* warranted. For there is something which is left unexplained, which is the nature and the source of the identity of concepts. We could ask: why do we have the capacity of redeploying our thoughts so that the very same concepts are entertained? And how do we know that they are the same?

As Peacocke mentions himself (p.150), a sceptic about all this could claim that when we redeploy our first-order contents in second-order beliefs, we do not need to rely on the latter’s *identity* with the first, but only upon their *similarity* with them. After all many views about propositional attitudes ascriptions rely on the view that such ascriptions

can be successful even when there is only an overall similarity between the concepts of the ascriber and the concepts of the ascribee. According to a well known view, we just project the contents of our own beliefs upon the other's beliefs, and this guarantees successful ascriptions in most cases. Why could it not be the same when we self-ascribe beliefs to ourselves?

But the considerations which precede seem to cast doubt upon this possibility. If the mechanism by which we ascribe beliefs to ourselves was such a mechanism of projection or of simulation analogous to the one that is said to operate in the ascriptions of beliefs to others, we would have to ascribe beliefs to ourselves in the same way as we ascribe beliefs to others, in the third person way. But this is impossible, for the mechanism of ascription by projection supposes that we have already an access to our own beliefs. So even the projective method of ascription, we need some sort of self access. If this self access were not guaranteed in some way, even our ascriptions to others could not be successful. So even on a similarity of content view, we would need privileged access.

I am aware of the fact that this constitutes only a form of a priori argument in support of the necessary entitlement thesis, and that it does not solve the problem of the nature of this privileged access. If, as I have granted, the necessary entailment thesis must be backed also by a causal account of our our beliefs can be available to ourselves, even when they are not necessarily conscious, this causal account has yet to be given. What I have claimed is that there are strong a priori grounds in favor of the necessary entitlement thesis. Since I admit that there may be causal factors as well, I do not want to exclude a priori the fact that a better account of these causal factors could in firm the a priori account given here. But at least I hope that it would confirm it.

## REFERENCES

- Armstrong, D. 1968 *A Materialist theory of the Mind*, Cambridge, Cambridge University Press
- Burge, T. 1988 "Individualism and Self-Knowledge", *Journal of Philosophy*, reprinted in Cassam 1994
- 1993 "Content preservation", *Philosophical Review*, 102, 457-488
- 1996 "Our entitlement to self knowledge", *Proceedings of the Aristotelian Society*, XCVI,91-116
- Cassam, C. ed, 1994 *Self-Knowledge*, Oxford, Oxford University Press
- Davidson, D. 1987 "Knowing one's own mind", *Proceedings of the American Philosophical Association*, 60, 441-58, reprinted in Cassam 1994
- Dennett, D. 1991 *Consciousness Explained*, Little Brown, New York
- Engel, P., 1996 *Philosophie et psychologie*, Paris Gallimard 1996
- to appear "Une conception matérialiste de l'esprit évacue-t-elle la subjectivité?", in *Cahiers de philosophie de l'Université de Caen*, 28, 1997.
- Heal, J. 1994 "Moore's paradox, a Wittgenstein solution", *Mind*
- Mellor, D. 1991 *Matters of Metaphysics*, Cambridge, Cambridge University Press
- Peacocke, C. 1996 "Our entitlement to self knowledge", *Proceedings of the Aristotelian Society* ,
- Rosenthal 1993 "Thinking that one thinks", in Davies and Humphreys, *Consciousness*, Blackwell, Oxford
- Ryle, G. 1949 *The concept of Mind*, Hutchinson, London
- Shoemaker, S. 1996 *The first person perspective and other Essays*, Cambridge, Cambridge University Press
- Wittgenstein, L. 1980 *Remarks on the philosophy of psychology* , vol. 2, ed. Blackwell, Oxford