

GIANFRANCO SOLDATI

SELF-KNOWLEDGE, REFERENTIAL IMMUNITY AND THE CONCEPT SELF*

1. Metaphysical idealism

I know that I was born in Locarno, that I am 80 kg heavy, that I am sitting on a chair, that I am writing an article, that I believe myself to be in Germany, that I have backache, and finally, I know who I am. In order to express these forms of knowledge I would refer to myself by using the term 'I'. I would say: 'I was born in Locarno', 'I have backache', etc. As such these forms of knowledge constitute *self-knowledge*¹.

I could know that the person who is writing right now is 80 kg heavy, that the person who is sitting at point x believes himself to be in Germany, and that G.S. is the son of so and so, without attributing any of these properties to myself. Although the latter cases of knowledge concern myself, they do not constitute any form of self-knowledge. They are cases of knowledge in the third person about myself.

What distinguishes self-knowledge from knowledge in the third person about myself? What sort of knowledge do I lack when I know that the person who speaks right now was born in Locarno without knowing that I was born in Locarno? Is there anything I have to learn in order to get from knowledge in the third person to self-knowledge? These questions are at the centre of the philosophical inquiry into the nature of self-knowledge.

Suppose I know a number of facts about an object a , such as the fact that a has a sister, that a owns a piano, etc. What other fact should I know in order to move from the above kinds of facts to the knowledge that I have a sister, etc.? A natural suggestion would be that I need to know that I am a . But an idealistically oriented philosopher might argue that there is no such fact for me to know.

* Earlier versions of this article were read at the SIFA-Conference 1996 in Vercelli, at a CREA-seminar organised by Gloria Origgi in Paris during the same year and at the first Tübingen-Konstanz meeting in 1997. Thanks to the participants for helpful comments. For specific comments I wish to thank Roberto Casati, Fabian Dorsch, Michele di Francesco, Frank Hofmann, Jürgen Pafel, Robert Pippin, François Recanati and Hans-Jörg Stotz. Special thanks to Manfred Frank who warned me against many popular as much as wrong views on self-knowledge.

¹ In an article which has inspired much of what follows, John Perry opts for the expression *self-belief* instead of the more familiar *self-knowledge* 'since the distinction between knowledge and mere belief is orthogonal to the issues I discuss' (J. Perry, 1990, p. 17). The same applies to most of the topics treated in this article. Nevertheless, since at some point it will be useful to assume that the subject really has knowledge about himself, I shall adopt the more traditional terminology. Notice, furthermore, that there are good reasons for distinguishing self-knowledge from self-consciousness (J.-P. Sartre, 1948 and M. Frank, 1991, p. 7). This paper exclusively treats of self-knowledge. Thus, at very least, it should provide a contribution for a better understanding of the background against which the question of self-consciousness ought to be raised.

There is nothing, he would maintain, which makes the identity statement 'I am x' true for any 'x' different from 'I'².

The argument need not involve any tendency towards mysticism. As in many other domains, the idealist receives the unexpected support from the eliminativist who believes that the expression 'I' does not refer at all³. There is nothing which makes the identity statement 'I am x' true because the expression 'I' simply does not refer. But this move jeopardises the phenomenon we started with: why worry about self-knowledge if there is no self to know anything about?

The idealist who is not willing to take the eliminativist shortcut should thus countenance the referential character of 'I'. He should admit that in 'I am F' the expression 'I' has a referent just as much as the nominal expression does in 'The person who is writing right now is F' or in 'G.S. is F'. But this concession has a dramatic metaphysical consequence: if the identity statement 'I am x' is never true in spite of the fact that 'I' has an authentic referent, this simply means that it is always false. There is no other expression denoting the referent of 'I'. The I, the subject, is an entity sui generis, categorially different from any other entity. Facts which are described by using the expression 'I' are irreducibly subjective. They cannot be described by any utterance not containing the expression 'I'. And they cannot be inferred from any premise not containing the expression 'I'. This is the position I suggest calling *metaphysical idealism*.

There are cases of self-knowledge where metaphysical idealism yields surprising consequences. There are a number of properties I have in common with ordinary objects such as tables, stones, etc. Take for instance my property of being 80 kg heavy. What makes the fact that I am 80 kg heavy categorially different from the fact that this table is, suppose, 80 kg heavy? The property of having a certain weight within a gravitational attraction field is typical of material objects with mass. If I have a weight, then I should be a material object as many others are. And there should thus be at least some identity statement of the form 'I am x' which is true.

The idealist who is willing to accept this argument will take a revisionist step. She will argue that the I who is 80 kg heavy is not the real subject and that my knowing that I am 80 kg heavy is not a case of genuine self-knowledge. Not all knowledge expressed in the first person is self-knowledge. In order to qualify as genuine self-knowledge, first person knowledge has to be acquired in a special way, typical of the first person. At this stage it is usual to refer to introspection: self-knowledge is introspective knowledge in the first person of facts concerning oneself. Introspection reveals only mental facts. Pains, desires and hopes are given to me in a way my weight, my place of birth and the colour of my hair cannot. Self-knowledge would thus reduce to introspective knowledge in the first person of mental facts concerning oneself.

² A position of this kind is described by Thomas Nagel: 'So what is left out of the centerless conception of the world – the supposed fact that I am TN – seems to be something for which there is no room in the world...' (Nagel 1986: 57). See also Nagel, 1970, p. 103. In the *Tractatus* Wittgenstein writes: 'The subject does not belong to the world, but it is a boundary of the world' (5.632); 'The philosophical I is not the human being psychology deals with, but the metaphysical subject, the boundary – not a part – of the world' (5.641).

³ Anscombe writes: «'I' is neither a name nor another kind of expression whose logical role it is to make a reference, *at all*.» (E.G.M. Anscombe, 1981, p. 32).

The notion of introspection has been severely criticised during this century. Some philosophers think that by itself, it is incoherent⁴. But the real trouble with introspection is that it has been charged with doubtful epistemological properties, such as infallibility, which has been attacked from various quarters over the last decades. The metaphysical idealist who adopts the notion of introspection, however, need not assume infallibility. All he needs is the claim that there is a special epistemic access to one's own mental facts which constitutes self-knowledge.

2. Semantic idealism

One does not need to accept the idea of irreducibly subjective facts in order to admit that the passage from knowledge in the third person to self-knowledge is far from trivial. It suffices to consider the following question: Is there any expression *x* different from 'I' with which 'I' in 'I am F' can be substituted such that a rational subject cannot believe that '*x* is F' without believing in the self-knowledge mode that he is F?⁵ If one's answer to this question is negative, if one believes that there is no expression satisfying this condition, then one believes that the concept <Self>⁶ is semantically irreducible. The concept <Self>, or <I> – as I will be occasionally calling it – would be a primitive, basic concept, which cannot be analysed with the help of any other concept. This is the position I suggest calling semantic idealism.

We have obtained the notion of semantic irreducibility by applying a version of Frege's Principle⁷: the expressions *x* and *y* have different meanings, are associated with different concepts, if a rational subject can accept a context where *x* occurs while rejecting the same context when *x* is substituted by *y*. Semantic irreducibility is the claim that there is no simple or complex expression having the same meaning as 'I'.

Semantic irreducibility does not imply metaphysical irreducibility. It is not a matter of denying the possibility that an identity statement of the form 'I am *x*' be true. Semantic irreducibility is the claim that such identities are epistemologically contingent, in the sense that for any rational subject it is always possible to take such a statement as false, whatever her other beliefs might be. There is no concept which warrants the passage from knowledge in the third person to self-

⁴ See for instance G. Ryle, 1949, p. 163-167.

⁵ When I say that *x* believes in the self-knowledge mode that he is F, I am attributing to *x* a belief he would express with *I am F*. Since this article doesn't deal in any specific way with problems related to belief attribution, I have not introduced any special notation for such cases. The context will do.

⁶ In the following I shall use simple angle brackets for conceptual contents. If *S* has a belief about the object denoted by *a*, I shall say that *S* associates to *a* the concept <*a*>, or that he *would express* the belief involving the concept <*a*> by using the expression *a*. If *S* has a belief he would express with (u) *a = b*, I shall say that he has a belief with the conceptual content <*a* is identical to *b*>. Notice that this is not the belief *S* would express with (u*) <*a*> is identical to <*b*>, which states the identity of the two concepts. The conceptual content of the belief expressed with (u*) would have to contain concepts of the concepts <*a*> and <*b*>. Since we shall never have to discuss cases of this type, there is no need to introduce any special notation for concepts of concepts.

⁷ See for instance: G. Evans (1982), pp. 18-19; S. Schiffer (1978), 180; (1990), p. 252; F. Recanati (1993), p. 75.

knowledge. Where metaphysical idealism postulates the existence of irreducibly subjective facts, semantic idealism maintains that there are irreducibly subjective thoughts. Semantic idealism, then, does not deny the equivalence of the fact that I am F with some fact described by an utterance of the form 'The ϕ is F'. It denies that the thought associated with 'I am F' can be identical to the thought associated with any utterance where 'I' is substituted by some other expression, different from 'I'.

Following metaphysical idealism the subject who has knowledge in the third person is bared access to a certain type of fact. Following semantic idealism he is bared from a certain perspective on facts. At this point it might be useful to note that where the metaphysical idealist can readily speak of missing knowledge pertaining to certain facts, the position of the semantic idealist requires some elaboration. He has to explain in what sense the acquisition of a new perspective amounts to an increase of knowledge. That said, however, it should be clear that contrary to the metaphysical idealist, the semantic idealist is not obliged to involve herself in any form of revisionism. Semantic irreducibility is independent of the kind of property the subject attributes to herself.

3. Epistemological idealism

There are those who think that self-knowledge involves access to irreducibly subjective facts and there are those who think that it involves grasping irreducibly subjective thoughts. Let us now consider the very common hypothesis that self-knowledge involves a particular form of knowledge, attainable only from the first-person point of view.

When speaking of the revisionist strategy of the metaphysical idealist, we already discussed introspection as a specific epistemic feature of self-knowledge. Yet, as we noticed at that stage, introspection obliges us to restrict the domain of self-knowledge to mental phenomena. By founding self-knowledge on introspection one loses grip on the question we started with. Instead of concentrating on the question of what distinguishes self-knowledge from knowledge in the third person, we are asked to inquire into the nature of the epistemic access a subject has to her own mental states. No doubt, this is an important question, but it remains unclear as to whether it gets to the heart of the problem of self-knowledge.

Let us ask, then, whether there is a specific epistemic feature of self-knowledge which does not involve such a restriction. Consider Shoemaker's immunity to error through misidentification relative to the first person pronoun. Shoemaker explains that:

[...] to say that a statement 'a is ?' is subject to error through misidentification relative to the term 'a', means that the following is possible: the speaker knows some particular thing to be ?, but makes the mistake of asserting 'a is ?' because, and only because, he mistakenly thinks that the thing he knows to be ? is what 'a' refers to' (Shoemaker 1984, p. 7-8).

It is easy to find utterances which admit this sort of error. I see somebody leaving the room, I believe that it is Ralph and I say: 'Ralph is leaving the room'. I know that somebody is leaving the room, but I am mistaken in my belief that it is Ralph. Or: I know that Quine wrote *Word & Object*, I believe that this man in

front of me is Quine, and I thus conclude: 'this man is the author of Word & Object'. Following Shoemaker this does not apply to such utterances as 'I feel pain' or 'I see a canary': I cannot falsely believe I feel pain just because I know that somebody, whom I mistakenly take to be myself, feels pain. One could thus submit that self-knowledge involves beliefs which are immune to error through misidentification relative to the first person pronoun.

As it stands this epistemological claim does not involve metaphysical idealism. It is compatible with the idea that there are no irreducibly subjective facts. Referential immunity (I shall thus abbreviate Shoemaker's notion) concerns the way in which a subject gains knowledge about certain facts, not the nature of the facts that knowledge is about.

It is important to distinguish referential immunity from the sort of epistemic immunity often attributed to introspection. My belief that I see a canary could be wrong if what I see is actually a robin, or if I am just hallucinating. The immunity which interests Shoemaker concerns the referent of the expression 'I' only. One could thus think that referential immunity is independent of the property the subject attributes to herself. We would then have found an epistemic feature common to all cases of self-knowledge with which we started. Let us call the position following which self-knowledge is referentially immune knowledge about oneself epistemological idealism.

There are two main problems with epistemological idealism. First of all, Shoemaker does not seem to think that referential immunity applies independently of the properties the subject is attributing to himself. His argument is based on Wittgenstein's well-known distinction between two uses of the expression 'I': the use-as-object, as in 'I have grown 10 cm' and the use-as-subject, as in 'I have tooth-ache'⁸. Wittgenstein himself notices that there is room for error relative to the identification of the person referred to with 'I' in the former case, but not in the latter. Shoemaker adopts Wittgenstein's argument in order to isolate the use of the expression 'I' which he thinks is constitutive of self-knowledge. He does not believe that all cases I mentioned at the outset are genuine forms of self-knowledge. Self-knowledge is given only when the pronoun 'I' is used-as-subject. Wittgenstein's examples, no less than Shoemaker's, suggest that referential immunity applies only when the subject attributes to himself mental properties. I could be wrong, for instance, in my belief that my hair is ruffled simply because I take my doppelgänger to be an image of myself in the mirror. Shoemaker's referential immunity would thus require a restriction of the domain of self-knowledge to mental facts to the same degree as introspection.

The second problem concerns the idea of using referential immunity as a specific trait of self-knowledge. There are referentially immune beliefs about oneself one would not express by using the pronoun 'I'. Imagine for instance that a subject observes someone sitting on a chair and thus judges 'this person is sitting'. This demonstrative judgement may of course be false, but it appears difficult to imagine that it could be false because, and only because, the subject judges that someone is sitting on the chair and makes the mistake of identifying that person with the person he refers to with 'this person'⁹. If this is true, it certainly remains so when the subject is demonstratively referring to herself as 'this per-

⁸ See L. Wittgenstein (1958), p. 67.

⁹ See S. Shoemaker (1994), p. 130.

son'. Yet the belief the subject would express with 'this person is sitting on the chair' does not qualify as a case of self-knowledge on our standards. It is thus possible to have referentially immune knowledge in the third person about oneself. Not every referentially immune belief about oneself qualifies as self-knowledge.

Again, one does not find in Shoemaker's use of referential immunity the idea that all referentially immune beliefs about oneself constitute self-knowledge. As we saw above, Shoemaker introduces referential immunity in order to isolate a certain usage of the expression 'I'. To say that only referentially immune beliefs a person would express by using the pronoun 'I' constitute self-knowledge is of course not to say that every referentially immune belief about oneself constitutes self-knowledge.

Referential immunity is neither necessary nor sufficient for self-knowledge. Yet this does not mean, as it will appear in what follows, that it does not play any role in the explanation of self-knowledge.

4. Subjectivity and consciousness

I have designated as idealistic all positions which claim that self-knowledge involves an irreducibly subjective factor. If you do not deny self-knowledge as such, and if you are prepared to consider it as a natural phenomenon in addition to many others, then idealism obliges you to acknowledge that a purely objective description of the world cannot be exhaustive. There is something missing. However, positions diverge as to what constitutes the missing factor. It is either an ontological category of facts, or a semantic category of conceptions of facts, or finally, a particular epistemic access to facts.

These different types of idealism give rise to different reactions. Metaphysical idealism, with its typical dualistic implications, won't satisfy the materialists. Fallibilists will be sceptical when faced with the immunity claim characterising epistemological idealism. And if one is weary of the notion of meaning, one will hardly be prepared to entangle oneself in semantic idealism. At any rate, these are not the fields in which I propose to fight the battle. I suggest, more modestly, an assessment of the different positions in the light of the phenomena with which we began. The metaphysical, just as much as the epistemological idealist, ends up by stipulating a redefinition of the very phenomena in question. They are both obliged to say that not all the cases I initially mentioned constitute genuine self-knowledge, which, they contend, applies only to knowledge of one's own mental states. Semantic idealism alone appears to be free from such revisionist requirements.

It is true that one often has to redefine a phenomenon before providing an explanation. But such a procedure needs to be backed by proper arguments. When we ask what distinguishes self-knowledge from knowledge in the third person, it is *prima facie* surprising to be told that we should rather consider the sort of property the subject attributes to herself. Furthermore, it is somewhat suspect to maintain that self-knowledge is directed at irreducibly subjective facts, or that it is endowed with a referential immunity, if one sets out from the hypothesis that we should restrict self-knowledge to knowledge of one's own mental states. The latter, no doubt, are not given to the subject in the same way as physical states are. Yet, have we not been asked to bar physical facts from self-knowledge just in order to rescue irreducibly subjective facts and referential immunity?

The relation between consciousness and self-knowledge appears to be particularly fragile precisely in respect to referential immunity. When I see my leg, touch my knee or hear the rumbling noise of my stomach, I perceive my body as any other material object. Beliefs founded on those perceptions do not possess Shoemaker's referential immunity. But I can also perceive my body 'from inside', proprioceptively. Gareth Evans has argued that beliefs based on bodily awareness possess referential immunity. He claims that utterances such as 'Someone's legs are crossed, but is it my legs that are crossed?' or 'Someone is hot and sticky, but is it I who am hot and sticky?' do not make sense, when based on proprioceptively gained knowledge¹⁰. The fact that my legs are crossed is arguably physical. It thus appears that proprioceptive beliefs constitute a form of referentially immune self-knowledge which is not directed on mental facts. Contrary to what was suggested by Wittgenstein's and Shoemaker's examples, epistemological idealism would not involve any restriction of self-knowledge to the domain of mental facts.

There are two main objections to Evans's analysis of proprioception. The first is that proprioceptively given facts are not physical, but mental. The second is that proprioceptive beliefs do not possess referential immunity. I shall not deal with the first objection here. Let me concentrate instead on the second.

5. Are proprioceptive beliefs referentially immune?

Could I not have proprioceptive beliefs concerning somebody else's body? Armstrong has indeed claimed that

[...] we can conceive being directly hooked-up, say by transmission of waves in some medium, to the body of another. In such a case we might become aware e.g. of the movement of another's limbs, in much the same sort of way that we become aware of the motion of our own limbs. (Armstrong 1984, p. 113)

If this were so, a proprioceptive belief could be false because, and only because, the subject mistakenly identifies himself with the subject to whom he is hooked-up. It would thus fail to qualify as referentially immune.

Extending Shoemaker's definition quoted under 3 to beliefs, one might say that a belief which would be expressed by an utterance of the form:

(1) a is ϕ

fails to be referentially immune if it depends on two beliefs which would be expressed by utterances of the form:

(2) b is ϕ

and:

(3) a is identical to b

so that the belief expressed by (1) could be false because, and only because the belief expressed by (3) is false (i.e.: although the belief expressed by (2) is true).

¹⁰ G. Evans (1982), pp. 220-221.

The question, of course, is what 'depends' means in this definition. It appears reasonable to suggest that it has to do with the acquisition of the belief under consideration. The belief $\langle a \text{ is } \phi \rangle$ fails to be referentially immune if the subject gained it through a process which involves an identification as (3). Referential immunity is a property a belief does not possess merely in virtue of its truth-conditional content. Two beliefs with the same truth-conditions can differ in their epistemic status, one being referentially immune and the other not. The beliefs expressed by 'I feel pain' and 'G.S. feels pain' are normally acquired in such a way that they differ with respect to referential immunity, although they have the same truth-conditions. If this were not possible, epistemic idealism would imply, contrary to what we saw above, metaphysical idealism.

Referential immunity depends on the possibility of there being a misidentification in the process leading to the acquisition of the belief under consideration. Following Shoemaker's definition, to say that the belief $\langle a \text{ is } \phi \rangle$ is referentially immune means that it could not be acquired in a way which would leave room to misidentification. Shoemaker himself contrasts this notion of 'absolute immunity' (Shoemaker, 1984, p. 8), which he thinks applies to the belief expressed by 'I feel pain', with what he calls 'circumstantial immunity' (*ibid.*), which he thinks applies to the belief expressed by 'I am facing a table'. This last statement does not have absolute immunity, 'for we can imagine circumstances in which someone might make this statement on the basis of having misidentified someone else (e.g., the person he sees in the mirror) as himself. But there will be no possibility of such a misidentification if one makes this statement on the basis of seeing a table in front of one in the ordinary way (without aid of mirrors, etc.)' (*ibid.*). Circumstantial immunity then, contrary to absolute immunity, depends on the way the belief has actually been acquired. A belief is circumstantially immune if it has been acquired in a way which leaves no room to misidentification. A belief does not cease to be circumstantially immune by the simple fact that it could have been acquired in a way which leaves room to misidentification.

We thus have two notions of referential immunity:

- RI.1 The belief $\langle a \text{ is } \phi \rangle$ is absolutely referentially immune if it could not be acquired in a way which leaves room for the misidentification of a .
- RI.2 The belief $\langle a \text{ is } \phi \rangle$ is circumstantially referentially immune if it has been acquired in a way which leaves no room for the misidentification of a .

Absolute immunity is clearly stronger than circumstantial immunity since it has the additional requirement that the belief in question can be acquired uniquely in a way generating referential immunity. This might well be the reason for restricting genuine self-knowledge to the self-attribution of mental properties. There is only one way, it might be argued, I can come to believe that I feel pain, namely by consciously feeling them. And this way generates referential immunity. The belief $\langle \text{I feel pain} \rangle$ is absolutely referentially immune. Proprioceptive beliefs, on the contrary, appear to have at most circumstantial referential immunity. A belief such as 'My legs are crossed' can certainly be acquired in a way which does not generate referential immunity.

Armstrong's counterexample poses a problem for absolute, not for circumstantial referential immunity. Even if one could be hooked-up to the body of another, the actual proprioceptive belief one gained in the normal way does not cease to be referentially immune in the circumstantial sense. Remember that we

are trying to determine an epistemic trait common to all, or at least to as many as possible, cases of self-knowledge we mentioned at the outset. More precisely, we are enquiring into the possibility of finding an epistemological feature which would not lead to a restriction of the domain of self-knowledge to the attribution of mental properties. Absolute immunity is of no help in this sense. So, it is circumstantial immunity we should be using.

Circumstantial immunity appears to correspond to what Evans describes as identification-freedom in the following passage:

What we should say is that a judgement is identification-free if it is based upon a way of knowing about objects such that it does not make sense for the subject to utter 'Something is ϕ , but is it a that is ϕ ?' when the first component expresses knowledge which the subject does not think he has, or may have, gained in any other way (Evans, 1982, p. 190).

This yields the following definition of identification freedom:

IF The belief $\langle a \text{ is } \phi \rangle$ is identification-free if it has been acquired in such a way that it makes sense for the subject to utter 'Something is ϕ , but is it a that is ϕ ?' only if he thinks that his senses are, or may be, malfunctioning.

In the light of this criterion a belief of the form (1) fails to be identification-free just in case the question 'Something is ϕ , but is it a that is ϕ ?' would make sense to the subject even if he did not believe that his senses are, or may be, malfunctioning¹¹. This certainly obtains when the belief under consideration is not referentially immune, in any of the senses defined above. But following IF the belief under consideration does not cease to be identification-free just because the subject can imagine having acquired it in a way which would allow that question to make sense. This is precisely the sense in which a belief of the form (1) has circumstantial, but no absolute immunity.

6. *The Unity of the Self*

We are still left with the fact that there are forms of self-knowledge which are gained in ways which generate referential immunity and forms of self-knowledge which are gained in ways which fail to do so. It appears possible, then, to wonder whether these two forms of self-knowledge really involve the same concept $\langle I \rangle$. A further argument is required in order to counter this doubt.

One such argument might be obtained by applying Evans' Generality Constraint to the concept $\langle \text{Self} \rangle$. François Recanati argues that one should distinguish between the 'buffer $\langle \text{Ego} \rangle$ ' and the concept one has of oneself¹². The buffer $\langle \text{Ego} \rangle$ serves as repository for information gained in a special way, namely in one of the ways which yield referential immunity (Recanati, 1993, p. 88). But Recanati thinks that the concept $\langle \text{Self} \rangle$ cannot be reduced to such a buffer. It

¹¹ For the belief expressed by (1) to be referentially immune it is not required that the subject explicitly believes that his senses are functioning normally. See G. Evans (1982), p. 190 (footnote n. 61).

¹² See F. Recanati (1993), p. 123. Recanati uses bold characters where I use angle brackets.

must be 'hospitable to any type of information' (p. 125), even to information which is not gained by proprioception or introspection. As Recanati puts it: 'one must be able to entertain not only the thought that <Ego> is F, where F is some predicate knowledge of whose instantiation can be gained in the special way relevant to the buffer <Ego>, but also the thought that <Ego> is G, where G is some predicate knowledge of whose instantiation cannot be gained in the relevant way' (p. 123). On Recanati's account, then, it is constitutive for the concept <I> that it is involved both in knowledge gained in a referentially immune way and in knowledge not gained in that way.

Recanati argues that the application of the Generality Constraint to I-thoughts is justified by the holistic character of thought: 'something is not a thought if it cannot be integrated to our general thought system' (p. 124). But holism comes in degrees. Surely, it is hard to imagine a subject who has the exclusive capacity to think one type of I-thought, say a thought of the type <I think>. To have the conceptual ability to entertain such a thought means to be capable both of applying the concept <thinking> to other subjects and to apply to oneself some other concept. But why should we extend such a requirement to concepts a subject applies to himself on the basis of knowledge gained in the 'non relevant way'? Does not the range of concepts a subject attributes to himself on the basis of referentially immune knowledge suffice to establish the possession of the concept <I>? We certainly want the subject to be able to link referentially immune knowledge to knowledge about himself accumulated by testimony or inferentially. But aren't we putting too much weight on the Generality Constraint if we use it in order to make this link part of the possession conditions of the concept <I>?

The question about the unity of the self, as distinguished from the identity conditions of persons, concerns the unity of the concept <Self>. If we want to be able to say that there is just one concept <Self> involved in the beliefs <I feel pain> and <I was born in Locarno>, we must provide a substantial account of what allows a subject to have beliefs about himself in the self-knowing mode which do not rely on an identification-free reference to oneself. Other concepts present an analogous difficulty. Take for instance the psychological concept <pain>. Most people would agree that you apply the concept <pain> to me on the basis of evidence I do not need to rely on when I apply it to myself. The same concept can be applied on the basis of very different criteria. What guarantees that under these conditions the concept <pain> I apply to myself is the same as the one you apply to me? We need a substantial account concerning the link between these two forms of application of mental concepts. Following a well-known line of argument both criteria of application are constitutive for the concept <pain>. It has been argued that one cannot possess the concept <pain> if one does not know how to apply it both in the first person and in the third person perspective¹³. This double perspective strategy parallels the idea mentioned above that it is constitutive for the concept <I> to be involved both in identification-free and in identification-dependent beliefs about oneself. But we have seen that for a justification of this approach, the Generality Constraint remains insufficient. Whatever one thinks about applying it to <pain>, the double perspective strategy does not seem to work with <I>.

Consider Ernst Mach's well-known example: 'Not long ago, after a tiring railway journey by night, and much fatigued, I got into an omnibus, just as

¹³ See, for instance, E. Tugendhat (1979), p. 89.

another gentleman appeared at the other end. 'What shabby pedagogue is that, that has just entered' thought I. It was myself; opposite me hung a large mirror. The physiognomy of my class, accordingly, was better known to me than my own' (Mach, 1897, p. 4). We might attribute to Mach the two beliefs he would have expressed with:

- (6) That man looks like a shabby pedagogue.
 (7) I look like a shabby pedagogue.

It is generally said that (7), contrary to (6), expresses self-knowledge. But in the light of what we previously said this needs to be properly justified. The belief associated with (7) is based on the external perception of one's own appearance. It is not referentially immune. The concept <I> involved in that belief, one could thus argue, is that of an object-body, not of a subject body. It would then be possible for Mach to ask himself whether the referent of 'I' in (7) is really the same as the referent of the first 'I' in the utterance expressing the referentially immune belief:

- (8) I see that I look like a shabby pedagogue.

Of course, we do not expect Mach to ask himself this question. As we have said above, it is implausible to assume that there are two different concepts <I> involved in the beliefs expressed by (7) and (8). But why? It sounds preposterous to rely on the claim that if Mach allows himself to indulge in this question he does not really possess the concept <I>.

How does Mach move from (6) to (7)? He certainly needs some further premise in addition to (6). Consider: 'I am that man'. This would certainly do the trick, but it does not correctly represent Mach's epistemological situation. We are precisely assuming that Mach does not know that he is the man he refers to with 'that man' while entering the omnibus. The same, we ought to assume, is true for any description Mach might apply on the basis of his observation to the person he sees in the mirror. So, for any ϕ he might apply to the person he sees in the mirror on the basis of his observation, Mach does not believe from the outset <I am the ϕ >. A conjecture would be that what Mach needs is some sort of direct, non-inferential connection between a belief in the first person about himself and a belief about the man he sees in the mirror. Only if Mach knows independently of what he observes that he has some specific property which, as he discovers, also applies to the man in the mirror, can he move from (6) to (7).

The best way to test this conjecture is to illustrate some of the ways Mach might in fact come to believe (7). Let us suppose, to begin with, that there are three persons on the bus, including Mach. He wrongly believes that there are four persons. We might then distinguish two separate, although in fact interwoven, cognitive performances. A first performance would consist in acquiring the belief that there are not four, but just three persons in the bus. The second would consist in applying the concept <Self> to one of those persons. Let us see how these two performances might be carried out. Here are some possible, by no means exclusive procedures.

First procedure. The first step involves Mach's normal recognitional capacities. The easiest and most probable way is that being familiar with mirrors, Mach soon rea-

lises that there is a mirror in the bus and that what he sees there is an image of somebody on the bus. He thus corrects his belief: there are not four, but three people on the bus. Now he still has to find out whose person's image he is seeing in the mirror. He might look around him and observe that none of the people he can see resembles the image in the mirror. Suppose the man in the mirror wears a dirty coat. Mach observes that neither of the other two people are wearing a dirty coat. He thus starts to look at himself. He looks at his own coat and, we might hasten to add, comes to the conclusion that he is wearing a dirty coat. But wait: exactly how did Mach come to believe <I am wearing a dirty coat>? Isn't he simply looking at a coat which appears closer to him than the coat belonging to the man over there (in the mirror image)? Mach needs the further premise <I am the man whose coat I am looking at>. Yet that is just the sort of belief we are trying to elucidate.

How, then, does Mach arrive at the belief <I am wearing a dirty coat>? It seems that this has something to do with the location of the coat, or more precisely, with the spatial relation between Mach and his coat. For instance, he might realise that the coat he is looking at happens to be at the place where he is. He believes <I am here> and <the dirty coat is here>. Given his general assumptions concerning the spatial relations between coats and the people wearing them, he concludes: <I am wearing a dirty coat>. And from there, together with the beliefs he has acquired about the other people in the bus, he can safely infer what he would express with (7).

Second procedure. Another way Mach might come to believe (7) is by using his own bodily movements. He has already recognised the mirror. Suppose he now raises his arm. He not only sees the man in the mirror raising his arm; he also knows on the basis of proprioception that he is raising his own arm. Now Mach is ready to make the relevant inference. He believes that there is just one man on the bus wearing a dirty coat. He sees that the same man is raising his arm. He believes, on the basis of proprioception, that he himself is raising his arm. He thus infers that he is that man. And he so concludes (7).

Third Procedure. Imagine finally that Boltzmann, one of the other two people on the bus, hearing Mach mumbling 'that man looks like a shabby pedagogue', exclaims 'you are that man!'. How does Mach come to believe, on the basis of Boltzmann's utterance, what he would express with (7)? Some procedure, such as the following, might apply. Mach hears Boltzmann's voice and believes he is telling the truth. He assumes that there is exactly one person Boltzmann is addressing. He recognises that that person must be at a specific place *p* where one distinctively hears Boltzmann's voice. He believes <I am here> and <*p* is here>. He thus infers that he is the addressee and that he, therefore, is the man whose image he sees in the mirror. Again, it is crucial for Mach to have the belief <I am here>. If he did not have that belief, all he could infer is that the person who is at *p* is a shabby pedagogue. Under the given circumstances this would qualify as a belief Mach has in the third person about himself.

As I emphasised from the outset, the procedures just described are by no ways exclusive. It is not claimed that there might not be any alternatives in the offing. But before searching further, it is worth pausing on what we obtained so far. The crucial belief in the first and in the third procedure is the self-locating belief <I am here>. Mach knows that this belief is true independently of what he

observes or hears. It is on the basis of his observation, however, that he gains the beliefs he would express with 'the man with the dirty coat is here', or 'the addressee of Boltzmann's utterance is here'. Being here, in this case, is the property which Mach knows to have independently of what he observes and which, as he discovers, also applies to the man in the mirror. Similar considerations can be made with respect to our second procedure. Mach does not need to observe his surroundings in order to gain the proprioceptive belief <I am raising my arm>. He observes, however, that the man in the mirror, and no one else in the bus, is raising his arm. Again, raising one's arm is the property which Mach knows to have independently of what he observes and which, he discovers, also applies to the man in the mirror.

Both, the belief <I am here> and the belief <I am raising my arm>, are gained in a way which generates referential immunity. Unless Mach believes that his senses are malfunctioning, it does not make sense for him to wonder: 'somebody is here, but is it me?'. Nor does it make sense for him to wonder: 'Someone is raising his arm, but is it me', if the belief <I am raising my arm> has been gained proprioceptively and if Mach does not think that his senses are, or might be, malfunctioning.

It is important to notice that the judgements involved in the three procedures are otherwise fallible¹⁴. Mach's belief <the coat is here> might be wrong. He might suffer a visual illusion. Or he might start from wrong assumptions concerning the spatial relations between coats and the people wearing them. His belief that the coat of the man whose image he sees in the mirror is dirty, might be false too. Or, in the second procedure, Mach might wrongly believe that he is raising his arm – he might suffer a proprioceptive illusion. And so he might falsely believe that he is a shabby pedagogue. The same applies to the third procedure, where Mach could be wrong in believing that the place where the addressee of Boltzmann's utterance is, is the place he would refer to with 'here'.

Some of these contingent beliefs are responsible for the fact that Mach's belief, which he would express with (7), is not referentially immune. Take Mach's belief: <Boltzmann's addressee is here>. This belief could be wrong, as it often happens when the speaker says: 'I am sorry, I was speaking to the person sitting next to you'. Surely, the place where the person next to me is sitting, as opposed to the place where I am sitting, is not here. So Mach's possibility of asking himself: 'somebody is a shabby pedagogue, but is it me?', might come from the fact that he wonders whether the place where the addressee is, is really the place he would refer to by 'here'. Or, in the second procedure, from the fact that he wonders whether he did not suffer a proprioceptive illusion.

But this fact, the fact that forms of self-knowledge as the one expressed in (7) are not referentially immune, would not imply that they involve a special concept <Self>. In our three procedures, the concept <Self> involved in the inference is introduced by a referentially immune belief. Leaving all the necessary supplementary premises aside, we could represent the core of Mach's inference in the three procedures described above as follows:

(9) That man is ϕ .

¹⁴ There might a qualm about the nature of the self-locating belief <I am here>. This latter belief, it might be argued, is infallible, not only referentially immune. Nothing depends on this point in what is at stake here.

- (10) I am ϕ .
 (11) I am that man.

ϕ is both, a property Mach attributes to the man in the mirror on the basis of his observation, and a property he attributes to himself independently of any observation. This latter attribution yields a referentially immune belief. The inferential step from (10) to (11) presupposes that the two corresponding beliefs contain the same concept $\langle I \rangle$. Thus, although the belief expressed by (11) might not be referentially immune, the concept $\langle I \rangle$ occurring in it must be the same as the concept $\langle I \rangle$ occurring in a referentially immune belief. Mach's second procedure shows, once again, how important it is for proprioceptively accessible facts to be accessible to normal perception. Mach would not arrive at his conclusion if this were not possible. The same applies to self-location in Mach's first and third procedure. He would not arrive at the belief $\langle I \text{ am the addressee} \rangle$, if the place to which he refers by use of the word 'here' could not be identical with the place he might think of as $\langle \text{the place where Blotzmann's addressee is} \rangle$.

These considerations suggest the following, provisional conclusions. We infer many of the things we know about ourselves in the first person from what we externally perceive and from what others tell us. The forms of self-knowledge we thus arrive at do not qualify as referentially immune. The inferences we use in order to arrive at those forms of self-knowledge, however, contain a referentially immune belief as a premise. In order for the inferences to be sound, then, the referentially immune premise and the inferred self-knowledge which lacks referential immunity must contain the same concept $\langle I \rangle$. It is only because we have a referentially immune access to some of the facts we infer from what we see or from what we are told, that perception and testimony can yield self-knowledge. If metaphysical idealism were right, if the facts we have self-knowledge about were accessible to the subject only, we could never learn what we are from what we perceive and from what we are told.

References

- Anscombe, G.E.M. (1981), «The First Person», in E. Anscombe, *The Collected Papers of G.E.M. Anscombe*, Metaphysics and the Philosophy of Mind, Blackwell, Oxford, vol. II, pp. 21-36; first appeared in: Guttenplan, S. (ed.) (1975), *Mind & Language*, Clarendon Press, Oxford, pp. 45-65.
- Armstrong, D.M. (1984), «Consciousness and Causality», in Armstrong, D. M. & Malcolm, N., *Consciousness and Causality*, Blackwell, Oxford, pp. 103-192.
- Evans, G. (1982), *The Varieties of Reference*, OUP, Oxford.
- Frank, M. (1991), *Selbstbewußtsein und Selbsterkenntnis*, Reclam, Stuttgart.
- Mach, E. (1897), *Analysis of Sensation*, Open Court, Chicago.
- Nagel, T. (1970), *The Possibility of Altruism*, Clarendon Press, Oxford.
- (1986), *The View from Nowhere*, Oxford University Press, Oxford.
- Perry, J. (1990), «Self-Notions. Logos: Philosophical Issues» in *Christian Perspective*, 11, pp. 17-31.
- Recanati, F. (1993), *Direct Reference. From Language to Thought*, Blackwell, Oxford.
- Ryle, G. (1949), *The Concept of Mind*, Hutchinson & Company, London.
- Sartre, J.-P. (1948), *Conscience de soi et Connaissance de soi*, Bulletin de la Société Française de Philosophie, XLII, pp. 49-91.

- Schiffer, S. (1978), *The Basis of Reference*, *Erkenntnis*, 13, pp. 171-206.
- (1990), *The Mode of Presentation Problem*, in Anderson, C. Anthony, J. Owens, (eds.), *Propositional Attitudes: The Role of Content in Logic, Language and Mind*, Stanford, CSLI Lecture Notes, pp. 249-268.
- Shoemaker, S. (1984), *Self-reference and Self-awareness*, in S. Shoemaker, *Identity, Cause and Mind*, CUP, Cambridge, pp. 6-18; first appeared in *Journal of Philosophy*, (65), 1968, pp. 555-567.
- (1994), *Introspection and the Self*, in Q. Cassam, *Self-Knowledge*, Oxford University Press, Oxford, pp. 118-139; first appeared in French, Peter A., Uehling, Theodore E. & Wettstein, Howard K. (eds.), *Studies in the Philosophy of Mind*, Midwest Studies in Philosophy, Minneapolis, vol. 10, pp. 101-120.
- Tugendhat, E. (1979), *Selbstbewußtsein und Selbstbestimmung*, Suhrkamp, Frankfurt a.M.
- Wittgenstein, L. (1958), *The Blue and the Brown Book*, OUP, Oxford.
- (1984), *Tractatus Logico-Philosophicus*, in Wittgenstein, Ludwig, *Werkausgabe* (Bd. 1), Suhrkamp, Frankfurt a. M., pp. 7-86.