LESSONS FROM SARTRE

FOR THE ANALYTIC PHILOSOPHY OF MIND

§1 Using Sartre

According to a well known account phenomenology and analytic philosophy have a common origin in the attempt to found and defend the objectivity of logic and philosophy against psychologism, a tradition of anti-pyschologism going back ultimately to Bernhard Bolzano. The respective founding fathers (Edmund Husserl and Gottlob Frege) differ in their methods and points of departure, so that – so the story is told (cf. Dummett 1988) – at last analytic philosophy was more successful in that language as intersubjectively shared turned out to be the better foundation of objectivity than the realm of pure phenomenology, where phenomenologists disagree and cannot establish an intersubjectively valid method of *eidetic reduction*. Analytic philosophy of mind also shares with phenomenology the fundamental interest in intentionality. Accounting for intentionality – in terms of propositional attitudes – turned out not only to be successful, but became (in the guise of functionalism) the very paradigm of the philosophy of mind and the cognitive sciences. What is missing in that philosophy of mind – as its main proponents like Jerry Fodor readily admit (cf. Fodor 1995) – is an account of consciousness *as experienced* by someone. Others in the analytic camp have offered theories of consciousness focusing on phenomenality and so called qualia (cf. Chalmers 1996). What is mostly and strikingly missing in these theories are (sub-)theories or models of the egological structures of consciousness (i.e. a theory of the subjectively experienced or theoretically to be assumed agents/egos in consciousness). There are mostly reflections on the use of the personal pronoun "I" and a undifferentiated notion of a/the "self". It is here, I think, that the analytic philosophy of mind should revisit phenomenology again.

The egological structures of consciousness have been a – or even *the* main – topic of Kantian, Idealistic and phenomenological theories of consciousness.

I have chosen Jean-Paul Sartre as my point of departure, since I believe that he has an advanced theory of these structures, and that some of his insights are congenial to theses in the analytic philosophy of mind. Sartre develops this theory in *The Transcedence of the Ego* (Sartre 1937), the introduction to *Being and Nothingness* (Sartre 1943) and his talk "*Self-Awareness and Self-Knowledge*" (Sartre 1948).

There are positive and negative lessons from Sartre:

- Taking up some of his ideas one may arrive at a better model of consciousness in the analytic philosophy of mind; representing some of his ideas within the language and the models of a functionalist theory of mind makes them more accessible and integrates them into the wider picture.
- Sartre, like any philosopher, errs at some points, I believe; but these errors may be instructive, especially in as much as they mirror some errors in some current theories of consciousness.

This paper, therefore, is not a piece of Sartre scholarship, but an attempt of a "friendly take-over" of some ideas I ascribe to Sartre into current models in the philosophy of mind.

§2 Ordinary Language and the Self

Talking of the self or an ego is often ridiculed by analytic philosophers by pointing out that sentences like

(1*) I came around and I brought (with me) my Self.

(2*) She visited Frank and my I was there, too.

are ungrammatical. They are ungrammatical, if they are, in the sense of running against the meaning of the expression involved, i.e. their common usage. This is, however, a very weak argument. The strangeness of (1*) might be accounted for by a proponent of a Self in noting

the inseparability of person and self, so that it is no more strange than

(3?) I came around and I brought (with me) my body.

This may not work for all constructions, (2*) may be an example of real deviance. Such deviance, nevertheless, does not show much. Starting from ordinary usage sentences like

(4) Near heavy bodies space is curved.

(5) All full explanation has to consider the colour of the quarks.

are nonsensical as well, since there is nothing, according to our pre-scientific understanding of space and before redefining the notion, against which it can be curved; and subatomic particles simply have no colours. Once it is conceded that scientific language may deviate from ordinary and pre-scientific usage there is no exception for the philosophy of mind. Maybe "the I", different sorts of "Egos" and "the Self" are theoretical posits. Given a background theory, sentences like

(6) The I unites experiences to present the Self to us.

may not sound strange any longer.

The deviance from ordinary usage may be considered a special problem for philosophy in as much as it is assumed to merely work with our intuitive understanding of ourselves and the world. Although this is partially right, this poses no real problem. On the one hand this complaint cannot be brought forward by analytic philosophers, who – especially in the cognitive sciences – stress the continuity of scientific and philosophical methods. On the other hand the problem may be due to the intricate character of the distinctions involved. There are plenty of other concepts and distinctions introduced by philosophers to re-construct our ordinary understanding of ourselves and our access to reality (e.g. the terminology of "possible world" semantics, the vocabulary of epistemic appraisal and confirmation, like "falsifiable", "simplicity", keeping "indirect" and "direct" duties apart – and so on).

§3 Self Denial in the Analytic Philosophy of Mind and in Sartre

Another criticism has focused on "the" Self as a supposed *object* we encounter in self-awareness. The deeper – even if philosophically somewhat shallow – reason that self-awareness is neglected as a topic by many analytic philosophers may just be that it is understood as being the awareness of a self as an object. If we exclude the possiblity that self-awareness might be more, given the dubious character of "the Self" as an object, self-awareness drops out of consideration as being a mere by-product (a secondary construction) of more interesting and fundamental mental events.

Nevertheless there is something to this criticism. Marvin Minsky (1985) sees the self as a construct: Thoughts are outputs of the cognitive systems, where several agencies, each of which doing only its job, work in the background being involved in perception, association, memory access and where several information states compete for the access to consciousness; some of the information states model control states that work on lower states; from these states a *self-image* of the system is built up; this construct is the self, seen as the agent who has the thoughts in question and who is responsible for the actions of the system; the self is not some *additional agent* inside you looking at the performance of the other agencies; the self is a *representation*; the self is ascribed properties that are essential to give the system's self-representation unity; so the self develops as a *narrative* in which language is used to describe an entity with coherent properties. Similar accounts of the self as (narrative) construction one can find in (Dennett 1991) and (Metzinger 1995).

Interestingly this opinion is not far of from Sartre's. The me is, for Sartre, a posited transcendent object (cf. Sartre 1937: 70, 76). The self – called "ego" by Sartre here – is something brought *before* consciousness, is an object and not that which is intentionally directed at this object. The self is "an object", not something active. The self is posited *as* the origin of acts and *as* their principle of unification:

> [C]onsciousness projects its own spontaneity into the ego-object in order to confer on the ego the creative power which is absolutely necessary to it. But this spontaneity, *represented*

> and *hypostatized* in an object, becomes a degraded and bastard spontaneity, […]. (Sartre 1937: 81)

So we may understand the Self as representing the whole "society of mind" (with all its processes and agencies) as a single agent. With the concept of "the Self" we represent the whole system/architecture. This is not wrong in as much as that system is us, and is acting. It is misleading in as much as we might start a search for that *agent* Self that is not among the agents of the mind. The Self is nevertheless *phenomenally real* and can be described in its features. The self represents the unification process within the cognitive system, including the occurrence of deliberate (verbal) control states. Other features of the Self may correspond to hidden cognitive agents, and so again the Self as construct is not inadequate. It is, therefore, misleading to say that by positing the Self we are victims of an illusion. Sartre may come close to this (see also Priest 2000: 124-26), Metzinger (1995) really claims this; but the mere fact that the Self is a representation does not make it a misrepresentation. If the Self is a representation of the whole cognitive system its referent really does what it is described as doing. Even our narrative of the Self re-enters memory and so influences our further acts. For the phenomenology and structural modelling of self-awareness it is indeed important to see that the Self as representation is not the agent of the act. *Here* a hypostatization would block the view on the pre-reflexive structures of consciousness and the *Ego.* The decisive point is to see the Self not as the agent in control but as a (narrative) construct.

Having thus downsized the Self one has to avoid overdoing the deconstruction. Overdoing the rejection of supposed entities in the vicinity of self-awareness loses the phenomenon itself. The crucial distinction that is often overlooked, and which is at the centre of my paper, is that between the Self and – at least one – I, which both have to be kept apart from the person that I am. Sartre clearly sees that there is a question of the *Ego* to be considered after having set aside the Me. The phenomena put several questions to us either as phenomenologists or cognitive scientists by the phenomena.

§4 A Short Phenomenology of Some Distinctions

Here are a couple of basic observations concerning my knowledge and experience of myself:

Phenomenon I

"I" is a singular term. Singular terms are used in statements to refer to objects which are said to have some property, which is referred to by the predicate (the general term):

(1) The table in lecture hall 3F is white

Statement (1) is true if one has identified by the description (or its meaning) an object and discerns (by the meaning of "( ) is white") that it has the corresponding property. Singular terms serve to identify objects. Identification need not be successful. "The headless horseman" is a singular term, but refers to nothing.

The meaning of "I" is usually given as "the one speaking". That seems reasonable: If somebody uses the term "I" we (the hearers) know that she is talking of herself. Can "I", however, be employed to characterize self-awareness? – It seems not. Self-awareness cannot have the structure of the following statement:

(2) I see a white table in lecture hall 3F.

The question of identifying the referent (i.e. the question generally associated with singular terms) does not arise: I need not identify myself for myself. I am immediately present to myself.

Furthermore there is no chance of misidentification here. I am present in my consciousness and there is no one else whom I could mistake for the referent of "I" or whom I could mistake for myself. Furthermore I have to know myself as the one who does the identification in every act of identifying – even if I am not doing this in inner speech (i.e. I am not using the pronoun "I) *I* have to be aware of the act of identifying. And to identify *myself* I have to know myself already!

These phenomena throw a bad light on a propositionalist theory (employing the pronoun "I" to account for the structure of self-awareness).

Phenomenon II

I am a person. I can refer to that person for example by the description "the one who is lecturing on December, 18th, in lecture hall 3F at 4 p.m.". The description refers to me and I know that. I can describe myself in several ways, but not all ways of referring to myself as a person are dependent on a description. Some famous anecdotes highlighting my peculiar knowledge of myself make this clear: Jon Perry follows with his trolley a sugar line in the supermarket to draw the responsible customer's attention to his defective sugar bag. After a while he recognizes that he himself has laid the sugar line with a defective sugar bag in his trolley (cf. Perry 1979). – How can one describe this case?

Jon Perry had at some time t (when he started his search) an opinion with respect to the customer looked for. At this time t Perry is *de facto*, although he does not know it, this very customer. Perry has at this time *de facto* a belief about himself, only he does not recognize this. At a later time t* Perry recognizes that he himself is the customer looked for. Now he still has *de facto* a belief about the customer, but additionally he now has a belief *de se* with respect to *himself* (in an emphatic sense of "himself" which points to the self-access to be explained here).

This phenomenon shows that there is a difference between beliefs/attitudes in which I am referred to by a description and such in which *I* know about *myself*.

Phenomenon III

"The I/the *Ego*" sounds peculiar, echoing philosophical traditions out of fashion. With the first phenomenon, however, we have already seen that to know about some objects involves knowing in some way about myself as the one who knows the objects. There is obviously in any conscious mental event – if we stick to individual mental acts for the moment – something that attributes that very act to itself as the thinking "thing". And this I is not a modifica-

tion that sometimes occurs, as the anecdotes in the second phenomenon may make you believe, but is present in *every* conscious episode. (The anecdote is telling by being an instance of misdescribing myself using a description although I am immediately given to myself without using a description.) Even if I am not engaged in inner speech (processing thoughts in public language), but looking absent mindedly out of the window – nevertheless I know that it is *me* who is looking out. I do not have to use the pronoun „I" for this, I am just having my thoughts. There is no question as to who is having these thoughts. I am immediately given to myself (I am "at"/"by" myself). There are mental events (e.g. in phonetic decoding) which are not conscious, but if some act is conscious I am present. In *this* sense human consciousness is self-awareness (knowing oneself as thinking) – whatever forms of consciousness there might be in the animal kingdom. It is not the case that we first have consciousness and then – in some additional act? – there comes self-awareness. Whatever I know of consciously I know as known by me. Whatever content I am thinking I know about me. Mental content is content for somebody. This somebody (the I) is (phenomenologically) the same whereas the content changes. Although the content or the scene before my eyes changes I am still there. We experience a continuous agent of thinking while the content varies. The I does not fall on the side of mental content (in the sense of the observed scene, the sentence thought etc.). The I might be the agent I experience within my mental acts as the one who does the thinking (the supposed actor of the acts of thinking). Is it not the case that *I* am thinking – and not that thinking happens to me?

Phenomenon IV

There is, however, a further distinction to be made with respect to the just mentioned role of the *Ego*. Sometimes, although the question does not arise whose acts are these, I am absorbed in whatever I am doing. I am absorbed in looking at the cat playing with the cork, or I am engrossed in what I am reading. Then – without any effort – immediately I can become aware that *I* am looking at the cat, that *I* am reading. Now I am *explicit* about the subject of the act,

no longer is it only the content I was absorbed in that is presented. Sartre himself (1937: 46-49) uses the example of reading or looking at a picture. (Sartre 1948: 42-45) expresses the phenomenon as being at the same time at myself (because of the pre-reflexive *cogito*) and detached from myself (since it is *only* a pre-reflexive *cogito*, reflecting breaking the immediacy to the object). This shift is almost imperceptible. It is not that I consciously *intend* now to focus on myself or set out to see who is doing the thinking. It just happens that from one moment to the next I realize my *Ego* as being the subject of my acts. If there is some reflection involved here, it does not take place as explicit reflecting by some of my *acts* on *another* of my *acts*. If this shift towards the I is a reflection, it has to be modelled in some other fashion.

Phenomenon V

We have to add a phenomenological remark on (some) representations: Suppose you hear a bear humming. By the humming we refer to the bear as its source. We represent the bear *as* humming. The humming sound represents the bear in some fashion (including pitch, frequency etc.). The humming *itself*, however, by pain of a vicious regress, is not represented "as" itself. To hear the humming is nothing besides or above the fact of having some representation. Expressed as a general observation:

(F) There are representations with respect to which it is the case that their being tokened is accompanied by a phenomenal quality.

By tokening such a representation some quality is given in consciousness.

Several distinctions have to be made in the light of these phenomena:

1. "the Self" is that vague complex of biography and biographical knowledge, discussed in §3, that together with some body defines an individual person; names and descriptions refer to that person as known by me and others; the Self falls on the side of *content* of conscious states.

2. "the I/the *Ego*" is my I that, although in fact related to an individual Self, contains the structural functions which are shared by conscious beings (e.g. in the acts of perception mentioned above); let us call it the *Ego* or the *functional I*; in the light of phenomenon IV we will have to distinguish two components here, depending on whether (2a) the focus is on the I itself, which need not include biographical knowledge and so is not the same as (1), or (2b) on the objects that I am aware of.
3. "the implicit I" is the functional correlate of the functional I within the realm of tacit knowledge or mental events that are *not* conscious, but nevertheless are processed (e.g. in memory or pre-conscious association) as being self-attributed states.
4. the set of conditions necessary for consciousness to be possible at all, to arise in the first place are not present in consciousness itself; in correlation to the talk of the *Ego* as present in consciousness one might talk of a "transcendental *Ego*" here, but this analogy to an agent as we know it from consciousness may be simply mistaken.[1]

A theory of the logical structure of my knowledge of myself (including the *de se*-theory of self-awareness introduced in the next but one paragraph) deals mainly with the functional I, i.e. phenomenon (2), and secondarily with its relations to the other instances. It does not deal primarily with biographies or the Self. The talk of a transcendental unity of consciousness has been transformed within cognitive science into the talk about the architecture of a cognitive system that may give rise to consciousness. Keep also in mind the fact (F) about representations.

Sartre's theory also distinguishes between the Self/Me as a biographical construct and the functions of self-awareness. His distinction between a pre-reflexive and a reflexive *cogito* may mirror the distinction between (2b) and (2a).

§5 Sartre's Conception of the Pre-Reflexive *Cogito*

Sartre in his way defends the thesis that consciousness cannot be separated from self-consciousness, as was alluded to phenomenologically in the preceding paragraph. It is in this context that his introduction of a pre-reflexive *cogito* is crucial. It is a necessary condition for being conscious of some object to be conscious of being conscious, since an unnoticed consciousness is an absurdity (cf. Sartre 1943: 18). Consciousness presents itself (to itself). This cannot be another intentional act on pains of a regress of presupposed or required acts of consciousness. Thus the accompanying consciousness of oneself is no additional act besides the intentional act, and it is not a reflexive act having the intentional act as object:

> [T]his consciousness of consciousness … is not *positional*, which is to say that consciousness is not for itself its own object. Its object is by nature outside of it, and that is why consciousness *posits* and *grasps* the object in the same act. (Sartre 1937: 40-41)

This pre-reflexive *cogito* is within one and the same act that is a conscious act presenting some intentional object, it is not within a reflective act having the intentional act itself as an object. Neither does it come *after* there being some intentional act already, nor is it vacuously present to be filled then with content. There is only the one (unified) conscious state representing an object in which I am also (non-positionally) aware of myself (cf. Sartre 1943: 21). My being conscious of myself does not fall *not* on the side of the content of my conscious acts. It is responsible both for the content being conscious for me, although I do not focus on me, and is the precondition for the reflexive *cogito*. In having, then, a reflexive *cogito,* I once again have a pre-reflexive *cogito* in order for the act of reflection to be a conscious act.

Note for the following paragraph that that I which we call pre-reflexive *cogito* is not an *object* of thought as long as it is active in accompanying other contents. It is related to but not phenomenally identical to the I brought into focus by reflection. The latter, in addition, has to be kept apart from the Self. The pre-reflexive *cogito* does not have itself as an object, so we may model it along the line of fact (F) as some peculiar representation that with its mere occurrence has its crucial features. Since the pre-reflexive *cogito* is no act, it cannot be phenome-

nologically brought into focus itself, although the immediacy of any conscious act may be claimed as evidence *for it*. Its characteristic is only given negatively, in terms of what it isn't. For a theory of self-awareness we need a working model. Here we turn to some help from theories developed within analytic philosophy of mind.

§6 *de se* Theories of Self-Awareness

Within the philosophy of mind we can distinguish between phenomenological and psychological theories. A psychological account, say functionalism, refers to the role the state has with respect to other states or the system's behaviour. Within such an explanation it might be important that it "is like something" to be in that state, but not all psychological accounts of some states require that is feels like something to be in such a state. A psychological theory need not account for (all) phenomenological features of mental states. Therefore one and the same psychological theory is compatible with different phenomenological descriptions. A complete functionalist theory of self-awareness comprises:

1. the identification of self-awareness by giving criteria for is being ascribed and by explaining it's causal role.
2. the specification of the format of representation of mental content, which explains it's inferential structure and it's causal efficacy.

One and the same answer to (1) can be coupled with different answers to (2). The non-propositionalist account of self-awareness discussed here (a *de se*-theory) is an answer to (2). The *de se*-theory, therefore, is at least in part a phenomenological theory. The basic alternative is a propositionalist account in which all states of self-awareness (including the states/aspects enabling self-awareness) have to be propositional if not also sentential.

De se-theories (in short: DST) were developed by Roderick Chisholm (1981) and David Lewis (1979). I will not explain their theories, but take a few of Chisholm's considerations as

a starting point for some systematic explorations. Both theories are embedded in peculiar ontologies that need not concern us here.

Roderick Chisholm puts the basic thesis of a *de se*-theory as follows (cf. 1981: 1):

(A1) There are attitudes which are not propositional

but self-attributions of properties.

The objects of these attitudes do not belong to their content, as §4 said, so that the content consists just of the properties the supposed object is considered to have:

(A1') (i) Some contents of attitudes are properties.

Instead of *propositional attitudes* DST speak of attitudes in a more general way. Propositional attitudes are secondary with respect to the basic non-propositional self-attributions. (A1) is the fundamental structural axiom of DST. It uses the two *relata*: properties and I (see (A1')(ii) below). The fundamental relation is the relation of self-attribution which involves direct self-reference. (A1') contradicts the thesis of the propositionalist who claims that the content of an attitude can be given only by a proposition or a sentence. In a proposition or sentence properties are ascribed, but the referent (or its description) is part of the content. According to (A1) the object of some attitudes is descriptionless and, therefore, contentless. This object is, according to Chisholm, the I:

(A1') (ii) The I does not belong in/to the content of some attitudes.

To be justified is the following thesis:

(T1) The primary form of reference is direct self-reference.

This thesis should be justified by defining the ordinary ways of referring (usage of statements, singular terms, beliefs, perceptions…) with the use of the concept of direct self-reference.

It has to be shown, thus, that the following generalizations are true:

(T2) The primary form of belief is the self-attribution of properties.

(T3) The I is the primary object of my attitudes.

These basic ideas are taken up here. Of course it has to be made clear *which* I is the one that is a relate in conscious acts, considering Sartre's distinction between a pre-reflexive and a reflexive *cogito*. Sartre and the DST seem to agree that the subject of consciousness does not belong to the side of the content. Whether the reflexive *cogito* has to be taken as propositional, as one may take it in Sartre, is not that clear. The pre-reflexive *cogito* certainly cannot be, on pains of the well known regresses – here Sartre and the DST agree. Furthermore the talk of "object" in the DST, say in (T3) should either not be taken in the sense in which Sartre denies that the pre-reflexive *cogito* is the object of a conscious act, in which case (T3) would be false for it, or the talk of "object" should be taken as in Sartre and then there will be a distinction between the reflexive I, for which something like (T3) holds, and the pre-reflexive *cogito*.

§7 A Synthesis of the Pre-reflexive *Cogito* with a *de se* Theory of Self-Awareness

De se theories and Sartre's conception share the crucial axiom that the I responsible for being also aware of myself in being aware of something else is not part of the content of my thought proper. Self-awareness – and thus any consciousness, since the two phenomena cannot be brought apart – has two components: my knowledge of myself (not to be understood as a second act) and my attitude (believing, wishing, seeing…) to some content.

In this paragraph I try to build a synthesis of Sartre's idea of a pre-reflexive *cogito*, the distinction with reflexive consciousness, and a *de se* model of representation. As a means of presentation I use symbols like "☺", "😐", "☹" and others, alluding to the *Language of Thought* hypothesis (Fodor 1975), that there is a medium of representation in the mind that can be understood in analogy to (public) languages and may be seen as the programming language of the mind. This thesis will only be used in a vague or general sense, since so it will be easier to understand the psychological reality of the fundamental relation of self-attribution used in DST. Not much is said about the inferential role of such an I-symbol within

a LOT-model of self-awareness. That these symbols are looking funny should not be confused with the serious intent of the presentation. The use of these symbols circumvents some problems with keeping the different *Egos* apart linguistically, and avoids using expressions that carry heavy connotations in the history of philosophy (like „transcendental *Ego*“ etc.).

Suppose there is a *Language of Thought (LOT)*, then there is also a chain of LOT symbols corresponding to thoughts not rendered in inner speech. Taking some pictograms and capitalization as representation of LOT-symbols we may have, for example,

(1) ☎RED

as the representation that a (specific) telephone is red.

The structures of the *Language of Thought* are the structures of intentionality. We refer to some property by using or tokening the corresponding LOT-symbol (or some symbol of ordinary language). Someone tokens a LOT-Symbol if he produces a token of it (in his brain or “belief box”). To refer to some property is nothing else than tokening the LOT-symbol. Using the LOT-model we can try to make the representational structure of non-propositional consciousness plausible. If self-awareness was propositional it would have to have the structure:

(2) A believes that p.

Believing would be a relation to a sentence or proposition p. Put thus, the difficulty is that with the believer a subject seems to be presupposed with respect to which we can ask whether it is aware of itself (cf. §4). If it is self-aware the propositional structure adds nothing. If it is not self-aware, self-awareness had to arise by believing some special sentences/propositions, taking believing as such as not involving self-awareness. Which sentence/proposition should be able to achieve that? Take a sentence like:

(3) I am F.

The meaning “the one who is speaking” secures by the use of the pronoun “I” self-reference which is pragmatically immediate with the tokening of (3). This self-reference can also have a

special functional role. The processing of "I" can be explanatory for behaviour. The combination of (3) and (2) in third person reports like

(4) A believes "I am F".

could be explanatory for A´s self-directed behaviour. What this functional role, however, has to do with *phenomenal* self-awareness is not clear. It seems to be an addition to (3). When speaking in the first person, one would say

(5) I believe I am F.

If (5) is the *relatum* of my belief it seems that I am (as the agent of the thought) opposite or besides (5). If (5) was the structure of my self-ascriptions it had to be made certain that "I" refers to me, and that both uses of "I" refer to the *same* entity. The relate of my believing, if (5) was the structure of my thought, would be (3) again. The pronoun "I" can secure infallible self-reference, but phenomenal self-awareness might not *arise* thus.

If we have to presuppose phenomenal self-awareness, the processing of "I" is not necessary, even if „I" has a special causal role. I am given to myself and directly attribute to myself (without a further act of self-reference) the property *F*. The content of such an ascription is the property only, as (A1) of the DST in §6 says.

Now it seems that even in such self-attribution I *refer* to myself, however immediately. I know myself. The representation of this self-reference cannot be a symbol of a natural language, which by its meaning allows it to identify a referent, since the meaning of the symbol looked for cannot be intersubjective, the supposed meaning being my self-apprehension of myself. Subjective meaning are a *contradictio in adjecto*. Even claiming that different subjective contents correspond to the public expression "I" does not help, since this content, because of it being content *for me*, had to be my self-apprehension, but this whole self-apprehension we were trying to explain by postulating the *processing* of the (meaning of the) pronoun "I". So we had a second self-representation as the content of a part of the first self-representation (by using "I") leading us into a vicious regress. The representation of my self-

reference can, therefore, have no meaning (as meaning is usually understood). Let us suppose instead that "😐" is the LOT-symbol of immediate self-reference (the I-symbol). Self-attributions have then the structure:

(6) 😐 F

where “F” either is a general term of a natural language or the LOT-representation of a property. “F” stands within the scope of "😐". (6) models an act of consciousness the content of which is *F*. So “😐” is not part of the content, it is the awareness of oneself that accompanies the awareness of some content. It is Sartre’s pre-reflexive *cogito*. The pre-reflexive *cogito* has the same role in Sartre’s theory as my unmediated knowledge of myself has in a *de se*-theory of awareness. The self-access given with Sartre’s pre-reflexive *cogito* and that given with tokening of “😐” is *part* of the one conscious state, not a further positional reflexive act.

Thinking (6) as a whole has a propositional structure, but this should not be confused with the claim that the content of the thought would be propositional. "😐" is not part of the content of my thought. If my self-apprehension consisted in representing “😐“ to myself there would be a difference between my processing of “😐“ (analogous to hearing a word) and my understanding the content of “😐“ (analogous to understanding the word). So we would have two processes taking place. There are not these two acts in my consciousness, neither do I *meet* a self-symbol or the like. Therefore my self-apprehension *is nothing else* than tokening "😐". Remember the fact (F). As content of my belief I only experience “F” or the property referred to by “F”. Between me and my self-reference intervenes no symbol. The symbol is not for me, I am it. In §4.III we said the I is not within the content. The I-symbol is not for me, but I am self-aware in virtue of tokening the I-symbol. "😐" is not perceived or apprehended from some point of view within me. The pre-reflexive *cogito* is not apprehended itself. "😐" does not “stand for” something, but with its tokening self-awareness is presented. That “😐“ is not part of mental content does not mean that “😐“ does not contribute to the inferential role that representation like (6) have. (6) taken entirely has sentential structure. A full-fledged LOT-

theory should be able to specify inferential roles accordingly. "😐" has by its syntax a causal role, as all LOT-symbols do. The DST tries to explain the structure of acts in which "😐" occurs and their relation to the other attitudes and attitude reports in natural language using *inter alia* the pronoun "I"; something I go not into here, see (Chisholm 1981) for details. The fact (F) for ordinary representations – that the appearance does not appear itself again, as Husserl said – can now be reduced to "😐" possessing this crucial feature; other representation behave according to fact (F) in as much as they are the content of some state introduced by the symbol "😐".

So we have a correspondence of our awareness with a LOT-sentence like

(7) 😐 SEE ☎ RED

↓‾ mode of the act ← (percept of) a red telephone

such that I am conscious of it

What this modelling does for Sartre's theory is giving it a working background theory, cashing in in terms of a semi-formal model the talk of a non-propositional pre-reflexive *cogito*. The LOT-hypothesis – and the funny looking symbols like "☺" – provide a model of mechanisms connecting the workings of a cognitive system with the occurrence of consciousness. What the appeal to Sartre's pre-reflexive *cogito* does for the DST is provide further backing for the claim that one has to comprehend the being aware of oneself as distinct from the contents of consciousness, as something not be thought of as in the (propositionalist) higher order model of self-awareness.

§8 Unity of Consciousness and Reflexive Assent

Given the basic features of DST this paragraph takes up related problems:

(a) Accounting for the ascent from pre-reflexive *cogito* to *presenting* an I to myself

(b) Accounting for the *unity* of consciousness on its different levels.

(ad a)

Sartre goes wrong, I believe, in identifying the object given in an self-presentation with the Me, and so finally rejecting the *epoché*; see Sartre's way of equating "I" and "Me" in (Sartre 1937). If there is an *Ego* apart from the Self/Me then after the *epoché* not all egological structures are gone in favour of Sartre's "pure field of consciousness". A problem of the *epoché* is that by cutting of the objects as real one turns from being at the objects to focussing on *act content* thus getting into a reflexive state easily. But then – in virtue of being in a reflexive state – there is this persisting I, its ubiquity being due to the *epoché*.

Even though what we experience in our self-awareness is ourselves as the individual we are, there is the distinction between the Self/Me and the self-representation of the agent of consciousness, since the assent to this self-representation is functionally distinct from object centred consciousness, and the operation of assent can be characterised generally without paying attention to any involvement of biographic knowledge (as would be distinctive of an involvement of the Self).

The reflexive assent should not be modelled simply in the traditional way as one act having as an object another act, as Sartre himself mostly does (Sartre 1937: 45, but maybe in contrast to Sartre 1948: 42, 85.). The LOT-hypothesis gives as the means to model the assent as the relation and modification of I-symbols.

"😑" works as an operator and has to be distinguished from a further LOT-symbol for me, say "☺", which can occur within the scope of "😑". Consider, for example, a reflexive thought having me not only as the agent of the thought, but also as an object; this objectification could be done by something like "☺". "☺" in fact is the reflected *cogito*. "☺" stands for the *Ego*, that arises with the almost imperceptible shift of focus mentioned in phenomenon IV in §4. With the tokening of "☺" we have the presentation of an I to ourselves. The thought has a structure like

(1) 😑 THINK ☺ SEE ☎RED

being the thought that it is me who sees that the telephone is red. We can model the shift from being absorbed into seeing the red telephone to being aware that it is me who sees the red telephone as the shift from

(2) 😐 SEE ☎RED

to (1). The operation that is responsible for the shift can be described as a rule:

(R1) Whenever "😐" is put into the scope of another "😐", then the left most "😐" within the scope is changed into "☺".

That only the left most "😐" is changed is necessary, since there is just *one Ego* and not a nesting of *Ego*s in consciousness, even if there are higher order thoughts like

(3) I believe that I want that I believe that dogs are green, but they just aren't.

As mentioned already we need another self-representation for mere self-representation, i.e. not as tokening either the pre-reflexive or the reflexive *cogito*. This self-representation is needed for such nested occurrences like in (3) and at the level of sub-doxastic processing in the cognitive system. We take "☹" as the corresponding symbol of the LOT. The LOT-rendering of (3) then becomes something like

(3) 😐 BELIEVE

(☺ WANT ☹ BELIEVE ALL:[🐄 → GREEN]) & NOT(ALL:[🐄 → GREEN])

where I have an explicit thought about me.[2]

"☺" is not the Self (as biographical construct), but the *Ego* experienced, although posited as a representation in the scope of "😐", as the agent of the acts, giving them unity. This objectification "☺" of "😐" has the function of *presenting* to me myself *focussed* as the subject of my acts. This function is independent of the biographical narrative surrounding the Self needed e.g. in claims of responsibility and understanding ourselves *as persons*.

"☺" and "😐" are *not the same*; thus, as Sartre says (1937: 44), the occurrence of the *Ego* is not due to the fact that one and the same entity – beneath the level of the whole cognitive system – is reflected *in itself*, as some Neo-Kantians claim.

(ad b)

The question of the unity of consciousness appears either as the question what unites some content into a consciousness of something or the question what unites several acts into a unified consciousness. The first question is the topic of Kant's theory of the transcendental unity of apperception or a theory of the conditions for consciousness to arise. The second question is closer to the role of the *Ego* within the conscious acts. Sartre denies that we need the *Ego* to unite consciousness, since the temporal structure of consciousness (including retention and protention) and the holism of mental content would suffice for that (cf. Priest 2000: 36-42); but this may seem questionable, since temporal or intentional unification seems to presuppose that there *are* several acts within *something* waiting to be unified. Given the DST, however, we can formulate a simple rule of unification of content:

(R2) 😐 F & 😐 G ↔ 😐(F & G)

This means that on some level of information processing a conjunction principle within the scope of "😐" applies. A similar rule may apply for "☺" and "☹". The rule is not a deep explanation of the unity of consciousness, but merely a description of an architectural constraint. On the other hand there is nothing in it that commits us to conclude from the fact that some *cogito* is responsible for unification that it is not the pre-reflexive *cogito* that is central for self-awareness.

§9 Where Do Higher Order Theories of Consciousness Go Wrong?

The DST model is not a higher order theory of consciousness (HOT) as they are widely held in the analytic philosophy of mind (cf. Carruthers 1996, Rosenkranz 1995), but it has some of its features. The *Ego* only appears after a modification of awareness that resembles reflection (see §8). This bringing the *Ego* into focus, nevertheless, was not modelled as involving propositions or even sentences of a natural language, as a HOT would have it.

Is Sartre's conception of self-awareness compatible with a propositionalist rejoinder to the DST?

There is one obvious point of reply for a HOT, which is also the most fundamental: A theory of the logical structure of knowing oneself has to keep – for the sake of the unity of a functionalist account of the mental – the connection between the functional I of awareness and the implicit I of mental processing. A propositionalist theory can do this more systematically than a *de se*-theory, since in the propositionalist theory both levels have the same logical format. The basic claim of the propositionalist (cf. Pylyshyn 1989) is:

(P) Any propositional attitude, any information processing, explicit or tacit, but cognitively penetrable, has the form: I (ATTITUDE) SENTENCE.

For example,

(1) I believe that it is Monday.

(2) I see that the audience is falling asleep.

etc.

The thesis that all conscious events are propositional is compatible with the claim that some contents of conscious acts are non-propositional representations (example: "I see this: ✈", in which a picture is following after the colon). Perceptual scenes can be embedded in sentential frames.

The "I" as LOT-symbol "😐" or as a symbol of a natural language has, according to the propositionalist, the meaning "that which is tokening this very sentence" and, therefore, is immune from failure of reference. It refers to the thinking person. This "I", still the propositionalist speaking, does *not* yield phenomenal awareness immediately. "😐" is not the representation for this. Fact (F) does *not* apply to "😐". Phenomenal self-awareness – even if it does not occur as explicit (inner) speaking – occurs only if *in the scope* (that is in the sentence within the structure defined by (P) an I-symbol is tokened (be it one of a natural language or a corresponding symbol of LOT like "😐"), like we had in (R1). For the propositionalist the

unity of the levels of mental processing requires that mental events on different levels (i.e. some of which are conscious, some of which are *not*) be within the scope of an I-symbol, whereas only those where an I-symbol gets into the scope of an I-operator yield self-awareness. What happens by bringing "☺" into the scope of "😐" is the *decisive* step from tacit processing to phenomenal self-awareness. This differs from the DST, where the mere presense of the pre-reflexive *cogito* (alias 😐) gave rise to awareness. Whereas DST is a "first order" theory (self-awareness arising by tokening a special symbol) the propositionalist account is a higher order theory (only by some representation being represented or being brought into the scope of another does self-awareness arise). The corresponding cognitive architectures or models of inferential roles might vary accordingly. Nevertheless the general idea of accounting for self-awareness by a process of tokening some LOT-symbol is kept also in the propositionalist theory. A radical version of a propositionalist theory could even claim that the I-symbol that matters is the pronoun "I" of a natural language. It helped build up the structures that matter for a functional architecture with consciousness. A less radical version could admit the secondary role of the pronoun "I", and might agree to denying a speaker meaning to "I", but would still see the structure (P) as the defining structure of self-awareness. Furthermore, the fundamental role which attitudes *de se* have according to Chisholm need not be denied, the propositionalist just sees this fundamental role for *de se* propositions. The only thing left over from DST then will be claim of direct attribution of properties. This claim was motivated by phenomenological considerations of how we know of ourselves within our states and as *not* being part of *the content* of the states which we experience. Can this phenomenology be undermined? Can the arguments given in §§4-7 be circumvented? In fact the justification given there depends on the analysis of the sentences

(3) I am F.

and

(4) I believe I am F.

It was claimed that these sentences cannot express the phenomenal content of self-awareness, since the agent believing these sentences would occur “on the other side” of this content. If these sentences are the content of my thoughts where am I? It seems that I am the one thinking the content, i.e. being related to the content and therefore distinguished from it. The analysis operates with a principle which could be expressed thus:

(E) That which is experiencing is not itself an experienced object in that act.

Now *suppose* it is the defining and peculiar characteristic of the I that it knows itself and *at the same time* is presented as part of the content of consciousness. The I-symbol then would instantiate my knowledge of myself and at the same time be part of the represented sentence. Why should it be impossible that I know myself as the continuous agent representing content and at the same time represent *that very agent* (not only myself in the manner of *another* representation like "☺") as that object to which some properties are attributed? This would have to be done by a *single* representation to avoid the problem of identifying the referents of the symbols. Self-awareness cannot arise by one I reflecting on *another*. The second I-symbol in (4) must not be a mere *objectivation* of the I, however that might be possible. The traditional opinion that subject and object are “one” or “united” here is a mere re-description of the problem. The traditional thesis (in Schelling or Natorp) that the acting I cannot be completely objectified leaves open to account for the mechanism of incompletely objectifying that very agent.

Phenomenologically it is not that clear as §4 made us believe whether self-awareness is non-propositional: Since I always am aware of myself when I am attributing myself – directly, since I do not have to identify myself – a property (like “😐 LookOutOfTheWindow”), this very knowledge has to be part of the content of what I am thinking. Where else should it be? What I know – even if it is knowledge of myself – seems to be mental content. If we put this knowledge into the processing of the I-symbol we are back at the propositional structure of (3) in §7! But putting it there is more than dubious for the reasons given in §7 and merely

saying that (E) might be false, as in the beginning of the preceding paragraph, does not give us a model of how this might be. For Sartre giving up (E) and thus going back to a propositionalist account in which the *cogito* in every case is part of the content is unacceptable; the pre-reflexive *cogito* is *defined* as being non-positional. It is thought of as a non-thetic consciousness, and thus cannot be modelled in the propositionalist fashion. Further on what would become of the shift between being absorbed in the content, although being conscious, and being aware that I am thinking these contents? This focussing on oneself simply does not seem to have the higher order reflexive structure the propositionalist assigns to it. Thus Sartre's theory of consciousness appears to be congenial to a DST account.

§10 Conclusion

One major shortcoming of the analytic philosophy of mind seems to be its inability to keep the Self as constructed biographical object sufficiently distinct from the *Ego* as the subject of our conscious acts. Even if the *Ego* is an aspect/is tied to a Self, its functions and its phenomenology require a theory of their own. Narrowing the attention to the Self downsizes self-awareness to an awareness of an object "Self". A motivation for avoiding a theory of the *Ego* may have been the fear of being committed to extravagant metaphysics. Keeping Self and *Ego* apart, however, allows one to substantiate the thesis that all awareness of something is at the same time awareness of oneself. Sartre's version of this thesis, using the pre-reflexive *cogito*, helps here. It can be synthesised with a *de se* account of self-awareness. Both parts may shed light on each other and come closer to saving the phenomena.

Notes

[1] One might – as Kant did – also speak of the transcendental synthesis or unity of apperception. I will not discuss this topic here. In Sartre it is clear that one should not confuse such conditions with the *Ego* as experienced by me. Sartre (1937) may be taken as accusing Husserl of confusing his talk of a transcendental *Ego* with Kant's talk of a transcendental *Ego*. I will neither discuss whether this interpretation of (Sartre 1937) is right nor whether Sartre himself represents Husserl's theory appropriately. Husserl (1913, 1931) is in his distinction between the empirical Me as a transcendent object and the *Ego,* which remains after the *epoché*, closer to the model advanced here. Husserl, however, takes that *Ego* as not being part of the content of acts, since he neither endorses a pre-reflexive *cogito* nor is he explicit as Sartre about the distinction between being absorbed in the intentional objects and focussing on oneself as having these intentional objects; cf. §8.

[2] This account of the phenomenality of myself experiencing myself is not that of the original DST in Chisholm (1981). Chisholm's theory works by a kind of "self-representing" properties. What these properties are and how they work seems to me to be part of Chisholm's arabesque ontology. The appeal to "self-representation" in properties either is only a title to the problem or has to appeal to something like (F). Since there are different *Ego*s to be co-ordinated, however, (see §4), we need also an account of their relation. An appeal to something like (F) is not enough at this crucial point of the theory. Chisholm also uses a relation of "considering" that one has such a property. This brings his account dangerously close to a higher order theory of self-awareness (see §9).

References and Further Reading

Carruthers, Peter (1996). *Language, thought and consciousness,* Cambridge.

Chalmers, David (1996). *The Conscious Mind*. New York/Oxford.

Chisholm, Roderick (1981). *The First Person*. An Essay on Reference and Intentionality. Minneapolis.

Dennett, Daniel (1991). *Consciousness Explained*. London.

Dummett, Michael (1988). *Ursprünge der analyischen Philosophie*. Frankfurt a.M.

Fodor, Jerry (1975). *The Language of Thought*. Cambridge/MA.

- (1994). "Jerry Fodor", in: Guttenplan, Samuel (Ed.) *Companion to the Philosophy of Mind*. Oxford, pp. 292-300.

Husserl, Edmund (1913). *Ideen zu einer reinen Phänomenologie und phänomenologischen Philosophie*. Tübingen.

- (1931). *Cartesianische Meditationen*; quoted by the Edition Hamburg, 1969.

Kenevan, Phyllis (1981). "Self-Consciousness and the Ego in the Philosophy of Sartre", in: Schilpp, P.A. (Ed). *The Philosophy of Jean-Paul Sartre*. La Salle.

Lewis, David (1979). „Attitudes *De Dicto* and *De Se*", *Philosophical Review*, 88, pp. 513-43.

McCulloch, Gergory (1994). *Using Sartre*. An Analytical Introduction to Early Sartrean Themes. London.

Merek, Prayoon (1988). *Sartre's Existentialism and Early Buddhism*. Bangkok.

Metzinger, Thomas (1995). "Faster Than Thought: Holism, Homogeneity and Temporal Coding", in: Metzinger, Thomas (Ed.) *Conscious Experience*. Lawrence, pp. 425-63.

Minsky, Marvin (1985). *The Society of Mind*. New York.

Natorp, Paul (1912). *Allgemeine Psychologie nach kritischer Methode*. Tübingen.

Perry, Jon (1979). „The Problem of the Essential Indexical", *Nous*, XIII, pp. 3-21

Priest, Stephen (2000). *The Subject in Question*. Sartre's Critique of Husserl in *the Transcendence of the Ego*. London.

Pylyshyn, Zenon (1989). *Computation and Cognition*, Cambridge, 5th Edition.

Rosenthal, David (1995). "Multiple Drafts and Facts of the Matter", in: Metzinger, Thomas

(Ed.) *Conscious Experience*. Lawrence, pp. 359-72.
Sartre, Jean Paul (1937). "La Transcendence de L'Ego: Esquisse d'une descriprion phénoménologique", *Recherches Philosophiques*, VI; quoted by the English edition *The Transcendence of the Ego*. New York, 1960.
- (1943). *L'etre et le néant*. Essai d' ontologie phénoménologique. Paris.
- (1948). "Conscience de soi et connaissance de soi", *Bulletin de la Société Francaise de Philosophie*, XLII; quoted by the German edition *Bewußtsein und Selbsterkenntnis*. Hamburg, 1973.

MANUEL BREMER