



Naturalizing Self-Consciousness*

Roberto Horácio de Sá Pereira**

Professor of Philosophy, Federal University Of Rio De Janeiro, Brazil

Abstract

The crucial problem of self-consciousness is how to account for knowing self-reference without launching into a regress or without presupposing self-consciousness rather than accounting for it (circle). In the literature we find two bottom-up proposals for solving the traditional problem: the postulation of nonconceptual forms of self-consciousness and the postulation of a pre-reflexive form of self-consciousness. However, none of them seems satisfactory for several reasons. In contrast, I believe that the only way of solving this traditional puzzle is to assume another bottom-up approach, namely the one that accepts Baker's challenge to naturalism and provides a naturalist framework for self-consciousness; in Baker's terms, to account for self-consciousness in non-intentional, non-semantic, and non-mental terms. That is the aim of this paper. My thesis rests on two claims. The first is the metaphysical claim that every creature enjoys a fundamental relation to itself, namely identity. The second is Dretske's epistemological claim that representations do not require a Self, traditionally understood as the principle that spontaneously organizes mental activity and lies behind all intentional acts. Briefly, I argue for a naturalization of self-consciousness that postulates non-linguistic, naturalized, and selfless form of representation of the cognitive system based on the metaphysical, fundamental relation everyone has to himself, namely identity. Self-consciousness emerges when brain states are selflessly recruited through learning to represent the cognitive system itself as a subject.

Key words: Self-consciousness; nonconceptual self-consciousness; pre-reflexive self-consciousness; naturalization of self-consciousness.

* Received date: 2018/07/15

Accepted date: 2018/08/26

** E-mail: obertohsp@gmail.com

Wo es war, soll das Ich werden

(Freud 1999: 16)

Introduction

The key feature of self-consciousness is *knowing* self-reference. What is in question is not the phenomenal character of what one *feels*, what it is like to be a Self, but rather how one *knowingly* refers to oneself. Given that the very act of self-reference must be self-conscious, the traditional puzzle of self-consciousness is how to account for such knowing self-reference without launching into an infinite regress or without presupposing self-consciousness rather than accounting for it (the vicious circle).

According to Kant (1956), self-consciousness in an *original* (ursprünglich), “*radical capacity*” (*Radikalvermögen*) (p. A114).¹ It deserves the label “transcendental” because: “[it] makes out of all possible appearances that can ever come together in one experience a connection of all of these representations in accordance with laws” (p. A108). And because: “...the synthetic unity of apperception is the highest point to which one must affix all use of the understanding, even the whole of logic and, after it, transcendental philosophy; indeed this faculty is the understanding itself” (p. B134). However, the most striking feature of Kantian self-consciousness is the fact that *its identity* is stated as a condition of cognition and hence is always presupposed by cognition and can never be explained in terms of more basic concepts or representations of pain of circularity: “I can ascribe them (representations) to the identical self as **my** representations, and thus can grasp them together, as synthetically combined in an apperception, through the general expression **I think.**” (p. B338. Original emphases)

In opposition to Kant, Fichte never took the identity of self-consciousness as “given”, but rather as something crying out for an explanation. However, he faces the following problem. As the identity in self-reference depends on an intentional act performed by the subject, she must be acquainted with herself as the subject performing the act *before* her own act of self-reference takes place. Thus, knowing self-reference is not accounted for but rather presupposed. In this regard, Fichte was certainly the first to formulate the puzzle of self-consciousness: The puzzle has nothing to do with the traditional Theory of Reflection and any attempt to solve it by means of some linguistic analysis misses the point completely.² Language only provides us with the fundamental data; the solution must come from a new view about the architecture of our cognitive system.

The puzzle forces us to face a dilemma. Either we accept the view that self-consciousness is the “highest point to which one must affix all

use of understanding,” or we dethrone self-consciousness of its transcendental condition and search for a bottom-up account for it. In contemporary philosophy we find two bottom-up approaches to self-consciousness. The well-known phenomenological solution to the puzzle consists of the postulation of a primitive pre-reflexive form of self-consciousness so that before carrying out the act of reflecting upon herself, the subject is already self-consciousness, albeit in an intransitive way. In contrast, Bermúdez’s solution consists of the postulation of primitive nonconceptual forms of self-consciousness. Both proposals have something in common with mine: they both undertake a bottom-up approach to self-consciousness. Yet, none of the proposals seems satisfactory to me. They raise several problems. I believe that the only way of solving this traditional puzzle is to assume another bottom-up approach, namely the one that accepts Baker’s challenge to naturalism (1998) and provides a naturalist framework for self-consciousness: to account for self-consciousness in *non-intentional, non-semantic, and non-mental terms*. That is the aim of this paper.³

My thesis rests on two claims. The first is the metaphysical claim that every creature enjoys a fundamental relation to itself, namely identity. The second is Dretske’s epistemological claim that representations do not require a Self, traditionally understood as the principle that spontaneously organizes mental activity and lies behind all intentional acts. Briefly, I argue for a naturalization of self-consciousness that postulates a non-linguistic, naturalized, and *selfless* form of representation of the cognitive system based on the metaphysical and fundamental relation everyone has to himself, namely identity. Self-consciousness emerges when brain states are *selflessly recruited* through learning to represent the cognitive system itself as a subject.

I shall proceed as follows. For those who are not acquainted with the technical meaning of self-consciousness, the first two sections are devoted to clarifying it. For those already acquainted with it, my advice is to skip them and go directly to the third section, which is devoted to presenting the traditional puzzle of self-consciousness. The fourth section briefly presents and criticizes Bermúdez’s account, while the fifth does so for the phenomenological one. In the sixth, I briefly explain how I understand self-consciousness and in the seventh I present my view of naturalization of self-consciousness. In the last, I briefly show why this self-consciousness is required.

Self-Consciousness: Linguistic Data

Self-consciousness in the sense of consciousness of oneself is a technical philosophical term that essentially means our ability to self-refer in a

particular way, namely, *knowing* that we are self-referencing (as opposed to a contingent form of self-reference in which the subject self-refers unknowingly). This same understanding is expressed by the usual characterization of self-consciousness as a reference to itself *as such* or by the expression reflexive or cognitive self-reference.

This immediate and conscious self-reference finds that the canonical expression in direct discourse (*oratio recta*) involves the use of the first-person pronoun: "I-thoughts" thoughts about oneself or, alternatively, "*de se* thoughts", the expression espoused by Lewis (1979):

1. I feel depressed.

And in indirect speech (*oratio obliqua*), the canonical expression of such knowing self-reference is the use of the indirect reflexive pronoun of the third person (a quasi-indicator) itself * as Castañeda suggests (1966):

2. I think that he* is depressed.

Some linguists, while independently exploring the presence in natural language of so-called *logophoric* expressions with more or less the same features that Castañeda ascribed to his theoretically posited quasi-indicators, have shown that PRO (the implicit subject of infinitive clauses that linguists posit following Chomsky) also appears to behave in English like a quasi-indicator.

2. I think that PRO to be depressed.

Of course, there are countless things that I know of myself, but that I do not know in this immediate way, for example:

3. I was born in Berlin, Germany, on 08/16/62.

The way in which I become aware of these many facts is not, in principle, different from the way a third person can know them: through observations and inferences. In fact, my parents have a greater epistemic authority than me regarding the truth of the proposition expressed by sentence 3: I know I was born in Berlin, Germany, on 16/08/62 because they told me. Following Tugendhat (1979), we can call the first form of self-consciousness expressed by sentence 1. *immediate* self-consciousness, while calling the second, expressed by sentence 3. *mediated* self-consciousness. The idea is to mark the fundamental distinction between a form of conscious self-reference that depends on the mediation of observations and inferences from another that is independent of these same observations and inferences.

However, the boundary between immediate and mediating forms of self-consciousness is fuzzy. There are innumerable properties that we can both self-attribute immediately, regardless of observations and inferences, but which third parties could equally assign to us with even greater authority. For example, let us say I think the following of myself:

4. I feel anxious.

Under normal circumstances, I know when I feel anxious without the need to observe myself or make any inference about myself. However, in observing me, my psychiatrist may know better than I do that the thought expressed by 4 is true. Besides, it is possible for a third person to know that I am anxious when I have not even realized it myself. Finally, it is possible for me to learn the truth of the thought expressed by 4, by observing my own behavior. Thus, I would have two ways of knowing the truth of the thought expressed by sentence 4: immediately and in a mediated way.

In fact, from the perspective of the third person I can come to know facts about myself without realizing that they concern me. The philosophical literature on self-consciousness is replete with examples in this regard. However, here we must distinguish cases in which this unconscious self-reference takes the *de dicto* form from cases in which it has the *de re* form. Castañeda (1966), for example, imagines a war hero; let us say General Bernard Montgomery, who after reading a detailed biography about himself has the following thought:

5. The commander of the Eighth British Army in the Second Great War in the North African Campaign was a hero.

However, being in an advanced state of amnesia he is not aware that by employing the description "the commander of the Eighth British Army in the Second Great War" he is self-referencing. Cases in which the reference is determined by the *satisfaction* of certain conditions of identification are denominated in the literature as cases whose reference is *de dicto*. In contrast, cases in which the reference is determined relationally are known as cases whose reference is *de re*. Perry (1979) provides another celebrated example of an unconscious *de dicto* self-reference. He imagines a guy in a supermarket following a trail of sugar on the floor, desperately looking for the shopper with the torn bag and thinking to himself:

6. The shopper with the torn bag is making a mess.

At every turn in his search for the shopper with the torn bag, the subject notes, however, that the trail becomes thicker and more conducive to the epiphany:

7. I am the shopper with the torn bag.

As in the previous case, by thinking about the thought expressed by sentence 6, the subject refers to himself without realizing it. This self-reference is *de dicto* because it is determined by the satisfaction of the condition of identification expressed by the description "the shopper with the torn bag" that is, by the *dictum* or proposition expressed by sentence 7. However, without knowing the truth of the proposition expressed by sentence 6 the subject nevertheless refers to himself because he employs

in thought a definite description, which in the context only he satisfies. But such self-reference is not conscious: it is merely contingent.

Perry (1979) gives us another celebrated example of non-conscious self-reference which is determined *relationally*. Perry portrays the experience of Ernst Mach, the famous physicist. After a tiring day at work, he entered a bus and saw a guy coming toward him through the window and thought:

8. But what a shabby professor!

It was only when Mach approached his seat that he had his epiphany: he was contemplating his own reflection in a large window:

9. I am the shabby pedagogue!

In the thought expressed by sentence 8 Mach's self-reference is determined by Mach's relation to the bus window and not by satisfying certain identification conditions as in 6 and 7. This means that the thought expressed by 8 is *de re*. However, even though the subject is Ernst Mach himself, it is not a *de se* thought since Mach ignores this fact.

As the *de se* attitude expressed by sentence 9 contains a direct reference, as in the *de re* attitude expressed by sentence 8, the imposing conclusion is that a *de se* attitude is nothing but a *de re* attitude whose *res* is the very subject. Still, rather than solving any problem, this linguistic analysis only provide us with the data that cries out for a cognitive explanation.

Self-Consciousness: Epistemic Data

It is important to make it clear that the immediate character of certain forms of self-consciousness does not mean infallibilism. Even without having to look at ourselves or making an inference about ourselves, we are often predicatively deceived as to the property we self-attribute, especially when the predicate is deferential. Thus, according to a famous example by Burge (1979), a patient is mistaken in reporting to his orthopedist:

10. I feel arthritis in my thigh.

No one can feel arthritis in their thigh because arthritis is an inflammation in the joints. Thus, by deference to the expertise of the orthopedist, the patient is allowed to correct as to the use of the predicate "arthritis."

Less often, we are also predicatively misled when we self-attribute non-deferential predicate concepts, that is, concepts that involve different levels of expertise about their linguistic domain. As we fall asleep, it is not uncommon to misrepresent ourselves kinesthetically as falling:

11. I'm falling out of bed.

Even more unusual, but not at all impossible, is error by identifying oneself in the case of immediate self-consciousness. Here it is not a matter of making a mistake by self-assigning a property that we do not possess. Since we possess the same property, we err because we assume that it is we who possess the property when in fact it is someone else. So, under normal conditions, I am not mistaken in self-referencing when I think:

12. I'm moving.

Who has never seen themselves in this situation? You are stopped in your car at a traffic light when you indirectly observe another moving car. The only thing your brain recognizes is that your car and the other are moving in relation to each other, but it does not know who is moving in relation to the ground, so you mistakenly suppose that it is you who are moving, which instantaneously causes you to brake.

Wittgenstein was certainly the first to realize this epistemic peculiarity by distinguishing two different uses of the first-person pronoun:

There are two different cases in the use of the word "I" (or "mine") that I could call "use as object" and "use as subject." An example of the first type of use is: "My arm is broken", examples of the second type are: "I see such-and-such," "I try to raise my arm," "I grow six inches," "I have a swelling on my forehead," "I think it's going to rain, "" I have toothache." (1958, pp. 66-67)

Examples of using the first person pronoun as an object:

13. I grew six inches.

Examples of the first, where "I" is used as a subject are "I see a computer in front of me" or:

14. I have toothache.

According to Wittgenstein, what would distinguish the use of the first-person pronoun as a subject from the use of the same pronoun as an object is the impossibility of a particular error:

One can point to the difference between these two categories by saying: Cases of the first category involve the recognition of a particular person and there is in these cases the possibility of an error, or as I prefer to say: The possibility of an error was established. . . . It is possible that, for example, in an accident, I feel a pain in my arm; I see a broken arm next to me and think that it is mine when really it is my neighbor's. And he could, looking at a mirror, confuse a swelling on his forehead with one on mine. On the other hand, there is no question of recognizing a person when I say I have toothache. Ask, "Are you sure you're in pain?" It would be absurd. (1958, p. 67. Original emphasis)

While the use of such a pronoun as an object involves the identification of the subject who self-attributes a particular property and is

therefore subject to identification-error, the use of the first-person pronoun as a subject would not be subject to the same type of error. However, the explanation that Wittgenstein gives us for such a distinction is entirely unsatisfactory. According to the Viennese philosopher, in using the pronoun "I" in sentence 14 I would not be making a judgment on a proposition that was true or false. Like nonverbal behavior, for Wittgenstein sentences like 14 would be mere expressions (*Äusserungen*) of mental states and not expressions of true or false propositions. Thus, there is no misidentification because there is self-reference in the first place.

Shoemaker assumes the distinction proposed by Wittgenstein, but argues that the use of the first-person pronoun as a subject is referential, albeit in a different way from its use as an object. It is easy to see why this is so. If it is possible to infer from my utterance of sentence 14 that someone (a bound variable) has toothache, then it seems undeniable that when I use the first person pronoun in 14, I make a judgment and self-refer. Shoemaker introduces the notion of immunity through misidentification relative to the use of the first person pronoun in the following terms:

To say that a statement "a is ϕ " is subject to error through identification with respect to the term "a" means that the following is possible: the speaker knows some particular thing to be ϕ , but commits the error of stating "a is ϕ " because and only because, he mistakenly thinks that the thing he knows to be ϕ is what "a" refers to. (1968, p. 557)

To understand what Shoemaker had in mind, it is useful to think of the inquiries we might make to someone who employs the first-person pronoun. When I use the pronoun of the first person *as an object*, it is theoretically possible that I am mistaken twice. First, a predicative error is never excluded. In the proposition expressed by sentence 13 it is possible that I was wrong to attribute to myself the property of having grown six inches. It is always possible that someone was wrong when measuring me or that someone deceived me to please me, because I am a child with problems of growth. Predictive error is theoretically possible when it comes to the question, "How do I know I actually grew six inches?"

But in propositions expressed by sentences in which the first-person pronoun is used as an object, it is also theoretically possible for me to be mistaken in terms of identification. So in the thought or proposition expressed by sentence 13, I could be mistaken as to the assumption that I am the one who grew six inches, when in fact it was someone else. Suppose my parents have the old habit of drawing marks on the wall to measure the growth of their three children. So, looking at a six-inch

difference between one mark and another, I might mistakenly believe that it was me who grew six inches when in fact it was one of my siblings.

Given this, what crucially distinguishes the use of the first-person pronoun *as an object* in sentences like 13 from the use of the same pronoun *as a subject* in sentences like 14, is the impossibility of misidentification. It is possible that I am wrong when I think it is toothache that I am feeling when in fact my mouth is completely anesthetized and what I feel is nothing more than the pressure of the shutter on my teeth. Still, assuming someone feels pain, it does not seem to make sense to wonder if I am the one who is really feeling pain and not a third person. And the reason given by Shoemaker is quite convincing: with the use of the first-person pronoun as a subject in sentence 14 there is no possibility of error by identification because such employment does not involve any identification in the first place (1968, p. 558). The characteristic mark of the "self" as a subject is precisely the absence of the identifying component.

Shoemaker conceives of two forms of immunity to error through misidentification. The first is what he calls *circumstantial immunity* manifested in propositions expressed by sentences like:

15. I'm facing a table.

In normal situations, when uttering 15 I may be mistaken as to whether the object before me is a table, but not as to the fact that it is me who finds myself facing such an object. There are circumstances, however, in which in asserting 15 I could erroneously mistake a third person for myself (when, for example, I made such a statement by looking at a mirror). According to Shoemaker, though, such circumstantial immunity would be derived from an absolute form of immunity to error through misidentification, which could be represented in terms of the following sentence:

16. I see a table in the center of my visual field.

If there are circumstances that could lead us to an error through identification when we think or utter 15, according to Shoemaker, there would be no circumstances that could lead us to an error through identification by reference to the use of the first-person pronoun when we think or utter sentence 16. Since the self-attribution expressed by 15 is seen as an inductive consequence of the self-attribution expressed by 16, according to Shoemaker, what confers circumstantial immunity to 15 is the absolute immunity expressed by 16. The contrast between 15 and 16 allows us to understand the origin of immunity to error through misidentification: while in sentence 15 the subject assumes a non-mental predicate (being in front of a table), in sentence 16 the subject self-attributes a mental predicate (perceiving a table).

Shoemaker's explanation of why the use of a certain sentence is immune to error through misidentification is wrong. As Evans correctly points out (1982, p. 220), immunity to error through misidentification is not restricted to the class of first-person psychological propositions in which a mental predicate is self-attributed. Suppose I think the following sentence:

17. My legs are crossed.

Under the assumption that I know someone's legs are crossed, it makes no sense for me to wonder if my legs are the ones that are crossed and not someone else's legs. But if this is so, the absence of identification in self-predication cannot be explained by the self-attribution of a mental predicate, as Shoemaker supposed. Evans's central thesis is that immunity to error by identification would not be a property of propositions *simpliciter*, but rather of judgments or beliefs about propositions from their modes of justification. That means that immunity to error through misidentification is not a semantic but rather an epistemological feature of self-consciousness. What delimits the class of judgments that are immune to error through misidentification from those who are not is the evidence from which they derive the information on which they are based. Thus, my judgment expressed by sentence 17 is immune to error through misidentification when its justification is based on proprioceptive sensations about the position of the members of my own body. In this case, the information that a particular body property is being instantiated is usually accompanied by the additional information that I am instantiating it. But the same judgment will not be immune to error through misidentification when my justification is based on my observation of my legs in a mirror.

Now, if it is the judgment expressed by sentence 17 that is immune to error through misidentification, when based on information obtained through proprioceptive channels, and not the proposition expressed by the sentence in question, then Shoemaker's distinction between circumstantial and absolute immunity to error through misidentification is untenable. In other words, there is no such thing as absolutely immunity to error through misidentification judgments.

Self-consciousness: Formulating the Puzzle

Fichte was undoubtedly the first to formulate the puzzle of self-consciousness, describing it as a "sophistry:"

We become (...) conscious of our consciousness of our consciousness only by making the latter a second time into an object; thereby obtaining consciousness of our consciousness, and so on *ad infinitum*. In this way, however, our consciousness is not explained, or

there is consequently no consciousness at all, if one assumes it to be a state of mind or an object and thus always presupposes a subject, but never finds it. The sophistry lies at the heart of all systems hitherto, including the Kantian. (*NI*, II, p. 356)

Henrich reformulates Fichte's paradox in the following terms:

(a) It is not difficult to see that the reflection theory is circular: if we assume that reflection is an activity performed by a subject – and this assumption is hard to avoid – it is clear that reflections presuppose an “I” that is capable of initiating activity spontaneously, for the “I” as a kind of quasi-act cannot become aware of its reflection only *after* the fact. It must *perform* the reflection and be conscious of what it does at the same time as it does it. (1971, p. 11)

Yet, Cramer certainly formulated the problem most clearly:

But how can the subject know the she in the reflection has herself as her own object? Apparently, only through the fact that the ego knows that she is identical with herself as her own object. Now, it is impossible to attribute this knowledge to reflection and to justify knowledge from it. *Because for every act of reflection it is presupposed that I am already acquainted with myself*, to know that the one with whom she is acquainted, when it takes herself as object, is identical to the one who is making the act of reflection turn back on itself. The theory, which wants to make the origin of self-consciousness understandable, therefore ends necessarily in a circle: that knowledge already must presuppose what it wants to explain in the first place. (1974, p. 563. Emphasis added)

Fichte's paradox can be reconstructed in the form of a classic dilemma. The first horn of the dilemma is the easiest way of formulating the puzzle. As we saw the distinguishing semantic feature of self-consciousness is the *knowing* self-reference expressed by sentences like:

1. I feel depressed,

in opposition to contingent self-reference, expressed by sentences like:

6. The shopper with the torn bag is making a mess.

Whenever I think sentence 1, I am acquainted with myself as the subject behind the intentional act of self-reference expressed by sentence 1. However, the question is: how can we account for this self-acquaintance? One way is to perform a second-order I-thought by means of which I identify myself as the author of the first-order thought 1:

18. I am the author of the thought expressed by sentence 1.

But in turn that requires the knowledge that I am now the author of the thought expressed by sentence 17:

19. I am the author of the thought expressed by sentence 18.

In this way an infinite regress is launched.

The second horn of the dilemma assumes that the knowing self-reference in the thought expressed by sentence 1 presupposes that in the thought expressed by sentence 1, the Self who is performing the act of self-reference already knows that she is the referred Self. In this way, claims Fichte, self-consciousness is presupposed rather than accounted for. This is what Fichte calls the vicious circle. Fichte's position is unclear, although very well known:

The "I" posits itself absolutely, that is, without any mediation. It is at the same time subject and object. The "I" only comes into being through its self-positing – it is not a preexisting substance – rather, its essence in positing is to posit itself, it is one and the same thing; consequently, it is immediately conscious of itself. (*Nl*, II, p. 357)

According to Henrich (1967), "Fichte has never explained his talk of positing and self-positing" (p. 18.). The formula "the 'I' posits itself" can only negatively characterize Fichte's own rejection of the need for self-identification. However, the idea of "self-positing" seems incomprehensible. Following this traditional reading, Fichte seems to mean that by self-positing, the "I" comes into existence. Yet, "how could someone perform that very act of positing if it does not yet exist in the first place" (Pothast, 1971, p. 71)?

Since Shoemaker's seminal paper, everybody recognizes that the only solution to the traditional problem is to assume that at the bottom level there is a self-reference that dispenses self-identification:

Self-knowledge might in some cases be grounded in some *other* identification, but the supposition that *every* item of self-knowledge rests on an identification leads to a vicious infinite regress. (1968, pp. 561-2)

According to Shoemaker, because we need to detain the regress, we must assume that in I-thoughts expressed by simple sentences like 1 there is no need for self-identification because there is no identification component in the first place. However, if we ask Shoemaker why sentences like 1 have no identification component he can only answer that in such sentences the "I" is used *as a subject* rather than *as an object*, that is, it is immune to error through misidentification. But that answer raises the question: Why are thoughts expressed by sentences like 1 identification-free? Shoemaker is certainly right, but he owes us a further explanation for that fact. To repeat Wittgenstein's jargon that the question about the need for identification in sentences like 1 is *meaningless* is an anachronism of the philosophy of ordinary language or of positivism that nobody else accepts.

The Putative Nonconceptual Self-consciousness

However, even if we assume that in sentence 1 with the first-pronoun used as a subject the question of self-identification is meaningless, the puzzle of knowing self-reference can be formulated in other terms. Bermúdez (1998) is the case in point. According to the deflationary view of self-consciousness, the ability to have first-person thoughts is reduced to the ability to employ the first-person pronoun in a way that reflects mastery of its semantics, that is, mastery of the token-reflexive rule of the first-person pronoun according to which the user of that pronoun knowingly refers to himself by virtue of his knowledge that he is the producer of that relevant token of the pronoun. In this view, the subject could not knowingly refer to himself in sentences like 1 unless he already knew that he was the producer of the token thought 1. Thus, to knowingly refer to himself in 1 *in a way that reflects mastery of semantics*, the subject is required to know beforehand that he is the producer of the relevant token of 1:

20. I am the producer of the token 1.

The obvious problem is that 20 is another *de se* thought that requires explanation in conformity with the subject's mastery of semantics, that is, in conformity with the same token-reflexive rule of the first-person pronoun. Moreover, the *de se* thought is not a clear case where self-reference dispenses self-identification. We are thus dealing with a circular explanation in the sense that we presuppose further *de se* thoughts (self-consciousness) rather than accounting for them. Bermúdez's way of defusing this "paradox" is to postulate nonconceptual forms of self-consciousness that are prior to and independent from conceptual forms of self-consciousness that rest on mastery of the first-person pronoun.

Bermúdez's nonconceptual solution to his paradox raises several questions that I cannot address here for a question of space. I confine myself only to two problems that directly concern the paper. To start with, Bermúdez's putative nonconceptual *de se* content does not seem sufficient to guarantee the truth of the *de se* thought expressed by 20. For one thing, being conscious in the proprioceptive way of one's own body and bodily limbs or being perceptually conscious of affordances etc. is not the same as *to judge* the proposition expressed by sentence 20. I may hear myself uttering sentence 1 or may read sentence 1 written down without recognizing myself as the real producer of the token of 1. I hear my own voice in the answering machine uttering 1 but fail to recognize myself as the producer of the relevant token and hence fail to recognize myself as the individual who is depressed at the time that token is produced.

Moreover, even assuming the existence of nonconceptual forms of self-consciousness to solve his paradox, Bermúdez has no solution to Fichte's puzzle. Insofar as for Bermúdez the nonconceptual forms of self-

consciousness take the form of self-perceptions, the subject must identify herself as the subject performing the act of perception and as the same subject perceived by that intentional act and hence she must already be self-conscious before she performs the act of perception. Again, self-consciousness is presupposed rather than explained.

Finally, Bermúdez mistakes nonconceptual contents for non-linguistic concepts. Nonconceptual contents are represented by mental states whose subject does not possess the concepts canonically required to specify what her states are representing. In contrast, non-linguistic concepts are those of creatures that do not possess a propositional language. When Bermúdez claims that the mastering of the token-reflexive rule of employment of the first-person pronoun must be accounted for at the nonconceptual level, he is identifying self-concepts with the mastering of the token-reflexive rule in question. However, a self-concept is a singular concept that each of us has of ourselves while the token-reflexive rule is general.

Given this, regardless of whether there are nonconceptual forms of self-consciousness, the simplest solution to Bermúdez's paradox is to assume the existence of non-linguistic self-concepts. The knowledge that I am the producer of the relevant token of the "I" results from my pre-linguistic self-consciousness based on my pre-linguistic self-concept.

Self-consciousness: the Phenomenological Solution

The well-known phenomenological solution to the puzzle consists of the postulation of a primitive pre-reflexive form of self-consciousness so that the subject is already conscious of her act of reflection, albeit in an intransitive or pre-reflexive way. There is no threat of infinite regress: because consciousness does not need a reflective consciousness to become conscious of itself. However, the postulation of a pre-reflexive self-consciousness is far from satisfactory and for many reasons. To start with, it is not clear how this pre-reflexive self-consciousness should be understood. In fact, some authors explicitly affirm that pre-reflexive self-consciousness can no longer be analyzed and must be taken as a fundamental piece of data. According to Frank: "However, we must also humbly declare that the basic element of our theory, familiarity, cannot be further analyzed." (Frank 2002, p. 400)

What we have is a purely negative characterization of the phenomenon. We know that self-consciousness should not be understood as a form of perception or reflection that takes the self (or some of its mental states) as its object, for this way of conceptualizing the phenomenon launches into an infinite regress. Furthermore, we know that self-consciousness should not be understood as objective consciousness.

On the contrary, I am pre-reflectively aware of myself only as “a subject.” Finally, we also know that such a pre-reflexive form of self-consciousness is neither mediated nor observational. The problem is how to characterize it positively. To assert that self-consciousness must rely on some kind of pre-reflexive and *immediate familiarity* with itself does not help us at all.

However, Zahavi seeks to positively characterize the idea of a pre-reflexive self-consciousness. In his words:

On the contrary, my pre-reflective access to myself in the first personal experience is immediate and unobservational and unobjective. It involves what has recently been called “self-reference without identification.” (2006, p. 280)

Zahavi’s reading is based on the analogy between Husserl’s and Sartre’s intransitive self-consciousness characteristics as non-observational and non-objectifying, and Shoemaker and Wittgenstein’s self-reference characterizations without identification as being Subjective (the use of “I” as subject” in contrast to the use of the “I” as object”). However, this terminological similarity is misleading. As Zahavi himself acknowledges, the phenomenological tradition conceives of pre-reflexive access to oneself in a non-referential way, and where there is not even self-reference, there cannot be self-reference without identification in the first place. In other words, while in Shoemaker’s explanation, the Self as subject is part of the representational content in the first person, the Self of Husserl and Sartre as subject is never part of any representational content.

Still, the main mistake lies in a false diagnosis of the puzzle. The puzzle is not a pseudo-problem that relies on the traditional Theory of Reflection, that is, on the assumption that in *de se* thoughts we are representing ourselves as one object among others in the world even when such representation does not involve self-identification. How could there be self-consciousness, that is, awareness of ourselves if we cannot represent ourselves as objects in the world?

In short, I see little point in the assumption of the existence of self-consciousness without reflexivity. The traditional problem of Fichte does not lie in the theory of reflection, but in the idea that in every representation of herself as an object the subject has to *identify* herself.

What does “Naturalizing Self-consciousness” mean?

I believe that the only way of solving the traditional puzzle is to accept Baker’s challenge to naturalism. In her words:

In this paper, I shall describe the first-person perspective, and then argue that philosophers and cognitive scientists have neglected the first-person perspective at their peril...I shall offer a challenge to naturalism: Either show how the first-person perspective can be understood

naturalistically, or show that it is dispensable. ...My aim is not to convince you that ... no science that aspires to be a complete science of everything can afford to ignore it. Thus, the first-person perspective is a good test case for naturalism. (1998, p. 327)

To show how a robust or reductive naturalism could accommodate the first-person perspective, one would have to give a nonintentional and nonsemantic account of the conditions under which an individual has that ability. I am not claiming that this cannot be done, but only that it is a challenge that robust naturalists have not taken up. (1998, p. 343)

In a sequence of ingenious papers (Backer 1998, 2012), Backer has questioned this widespread view that self-consciousness does not represent a problem of a naturalist program. Against the simple view, she argues, first, that self-consciousness, or what she calls the strong or robust first-person perspective, is irreducible to simple consciousness, which she calls the weak or rudimentary first-person perspective. As she understands the opposition, self-consciousness requires the self-attribution of the weak first-person perspective

According to Baker, the weak or rudimentary first-person perspective concerns thoughts expressed by sentences like:

21. I am happy (2012, p. 23).

While the strong or robust first-person perspective concerns thoughts expressed by sentences like:

22. I am glad that I am happy (2012, p. 23).⁴

However, it is questionable whether the content expressed by sentence 21 is not already a form of fully-fledged self-conscious thought. Indeed, the only difference between the thoughts expressed by sentences 21 and 22 is the fact that in 21 we have a first-order first-personal thought, while in 23 we have a second-order first-personal thought. To assume that self-consciousness entails the self-attribution of some first-order first-personal thought is too demanding. That is to over-intellectualize the phenomenon of self-consciousness. To be sure, I take Baker's side when she claims that self-consciousness requires knowing self-reference that is not present in simple consciousness. Whenever a newborn is colic, there is the feeling of what it is like to be colic, but there is no empirical evidence to assume that the newborn is knowingly self-referring.

Still, there is no empirical evidence whatsoever to assume that self-consciousness requires the self-attribution of knowing self-reference. When a mother asks her four years old child how she is feeling at her birthday party, she may answer by uttering sentences like 22. But there is no empirical evidence whatsoever that this child could mentally or linguistically articulate the thought expressed by sentence 23. The content expressed by sentence 23 requires much more than simple self-reference.

It requires some knowledge about how hard it is to be happy and expectations about one's own future. It seems to me to be quite plausible to assume that small children can possess the mental ability to knowingly self-refer by articulating first-personal thoughts expressed by sentence like 22 without having the higher-order ability of self-ascribing their first-personal thoughts. Be that as it may, this paper is not concerned with this issue.

To be sure, as Baker recognizes (1998), "naturalism" is an umbrella term that covers quite different things (p. 343). Still, she left it quite clear that naturalism cannot be understood as the mere refusal to appeal to something supernatural or immaterial. Yet, Baker is satisfied with this disambiguation and does not realize that her challenge might involve further ambiguities. First, the challenge is addressed both to philosophers and to scientists. Metzinger, for example, believed that he could deal with this challenge scientifically by elaborating a sophisticated neurological Self-Model that at the end of the day reveals self-consciousness to be an illusion (See Metzinger, 2003).

As Baker considers Metzinger's scientific answer to her challenge as an example of a "reductive naturalist account" (2012, p. 204), we are entitled to assume that her challenge is primarily addressed to cognitive science rather than naturalist philosophy. Still, I must make it clear from the outset that I see little point in understanding the challenge as addressed to science. To be sure, in my naturalist view of metaphysics there is a continuum between metaphysics and science. Many philosophical problems arise precisely because of apparent tensions or conflicts generated within the empirical sciences. Self-consciousness is one of many examples. To be sure, self-consciousness has always been an issue in the philosophical traditions at least since Descartes. Still, it is the recent development of cognitive science (basically neuroscience and cognitive psychology) that raises the question about the possible naturalization of self-consciousness.

However, I do not believe that we can solve philosophical problems scientifically. Even naturalized, philosophy of mind remains an *armchair activity*. Its job is to provide a general framework that must be filled with the empirical findings coming from cognitive science in general. If you believe that the naturalization of self-consciousness constitutes a philosophical problem, it is meaningless to attempt to solve it by looking for the neuronal signature of self-consciousness. A neo-dualist can easily concede that self-consciousness is not a separate substance from the body as Descartes famously claimed. However, by assuming that self-consciousness weakly supervenes on the brain, the anti-naturalist can hold on to his claim that self-consciousness is irreducible to its neuronal basis.

A third ambiguity concerns the indispensability of the first-personal perspective or of the essentiality of the indexical "I":

Two different sorts of considerations suggest that the first-person perspective is indispensable for our theorizing about reality. The first sort (I) concerns language: First-person reference is not eliminable from "I*" sentences, whether it is eliminable from simple, direct-discourse "I" sentences or not. The second sort (II) concerns psychological explanation: Certain psychological explanations of behavior require attribution of a first-person perspective to the one whose behavior is to be explained. (Baker, 1998, p. 24)

In those words, Baker suggests that the main obstacle to the naturalization of self-consciousness is Perry's celebrated claim that the first-person pronoun is an essential indexical that could not be eliminated either in favor of some definite description or in favor of some other linguistic self-referential device on the pain of leaving self-directed actions incomprehensible. As Perry (1979) has shown in several papers by means of different examples, we can only account for self-directed actions by assuming that the agent knowingly self-refers by means of the employment of the first-person pronoun. Thus, if Perry is right, the first-personal pronoun cannot be replaced by some definite description satisfied by the subject or by any other demonstrative-like self-referential device on pain of losing the capacity to make sense of the agent's self-directed actions. Again, Perry's thesis on the essential indexicals is just a piece of linguistic *data* waiting for explanation in terms of cognitive architecture.

Again, this is not the way that I find that accepting Baker's challenge to naturalism can solve our puzzle. As I wish to understand Baker's challenge, the naturalization of self-consciousness is not the challenge of elimination of the first-person pronoun in favor of other self-referential linguistic devices, or endorsing Metzinger's claims that self-consciousness is a mere illusion. Of the few things we know for certain, one of them is the existence of self-consciousness and consciousness in general.

The last ambiguity concerns Baker's opposition between the first-personal and the third-personal perspective. According to her: "the first-person perspective is a challenge to naturalism. Naturalistic theories are relentlessly third-personal" (2012, p. 203). In those terms, however, Baker restates the problem of the explanatory gap concerning all kinds of consciousness between processes in the brain and the ways that things taste, look, feel, smell, and sound: how can we explain that subjectivity emerges from the gray matter of the brain? If Baker aims to show that the naturalization of self-consciousness is a problem that is irreducible to hard

problems raised by consciousness *simpliciter*, this is certainly an unfortunate way to phrase her challenge to naturalism.

Now, let me explain in what sense accepting Baker's challenge can solve our puzzle. I accept her assuming some Kant-like viewpoint that self-consciousness is primitive in the sense that it cannot be accounted for in more primitive terms on pain of launching into an infinite regress or presupposing what is at issue. By accepting the challenge to naturalism my aim is to show how genuine self-consciousness ontogenetically emerges from a *non-mental, non-semantic, and non-intentional basis*.

Naturalizing Self-consciousness

As I said in the introduction, my thesis rests on two claims. The first is the metaphysical claim that every creature enjoys a fundamental relation to itself, namely identity. The second is Dretske's epistemological claim that representations do not require a Self, traditionally understood as the principle that spontaneously organizes mental activity and lies behind all intentional acts. The starting point is Dretske's naturalization of representational content. Firstly, we must understand the notion of information in a technical way: a signal (in our case, neuronal state) conveys information about a source (external object) when there is a nomic co-variation between them so this gives support to the following counterfactual idea: The signal would not occur without the source occurring. In probabilistic terms, given the occurrence of the source the probability of occurrence of the signal is equal to 1. But the registering of information is not enough to account for the concept of representational content, because it makes no sense to talk about *misinformation* in the technical sense. In contrast, there can only be representation where there is the possibility of misrepresentation. It is for this reason that Dretske must appeal to teleology. To acquire a representational content, a neural state (sign) that already conveys information about a source must be *recruited* with the function of indicating a given source. Dretske calls this *indicator-function*.

The fundamental distinction here lies between sensible representations of a non-conceptual nature and representations of a conceptual nature. In the case of non-conceptual sensorial representations, brain states are phylogenetically recruited with some indicator-function because of the adaptation of the species to its environment. Thus, for example, suppose that a certain neuronal state in my secondary visual cortex nominally covariates with the presence of a token of red, that is, it provides information about the presence of that token. However, since the ability to discriminate the color red from the others was of paramount importance for the adaptation of my species to its environment, the

assumption is that the neuronal states that convey this information about the token of red have been phylogenetically recruited by selection to indicate or represent that token of red.

In contrast, in conceptual representations, this indicator-function is ontogenetically acquired through learning. I only become able to represent the color red conceptually when further brain states in tertiary cortexes acquire a further indicator-function through learning: to represent all different tokens of red *as type*. Usually conceptual representations are based on previous non-conceptual representations. When this happens, Dretske used to say that the available information coded in analog form is transcoded into information in digital form, for example, all amounts of information about *tokens* of red is transcoded into information of the same red *as type*. But that is not mandatory: one's singular concept of God does not depend on any nonconceptual representation of him. Theoretical concepts of science also refer or fail to refer to independently of nonconceptual representation. In those cases we must assume that neuronal states are recruited with the function of indicating their reference solely through learning.

However, the most striking feature of Dretske's naturalistic and representational account has not received due attention. If Dretske is right, all modern tradition is wrong (with the exception of Hume) because initial representations are not dependent on some *subject* of representation. To begin with, nonconceptual representations are *selfless* in the relevant sense that they do not require a subject behind the representations. It is not because of our attentional efforts that our brains come to represent tokens of red nonconceptually. Likewise, the most primitive concepts such as the self-concept are not the product of any spontaneous thinking. It is not because of the spontaneous act of Reflection, that is, comparing, separating, and abstracting, that my brain comes to represent all tokens of red as belonging to the same RED color type. The vast majority of our brain states are *recruited* by natural selection or through learning/training with the function of representing something. Now, this is all we need to break Fichte's vicious circle or to detain his vicious infinite regress: it is not the Self as the spontaneous active principle that is behind self-reflection. Instead, brain states of the cognitive system are recruited *selflessly* with the function of representing the cognitive system.

The remaining question is how the subject comes to *know* that this self-concept refers to her Self in a way that explains the immunity through misidentification. As we saw, according to the tradition, this knowing self-reference depends on some *epistemic* self-identification. Here one mistakes one metaphysical thesis for another epistemological thesis. The ontological status of the Self is one of the most controversial questions in

philosophy. For Cartesians, the Self is an immaterial and immortal substance. For Kant it is just the faculty or agency behind the thinking creature. For Nietzsche, it is nothing but the living body. For Hume and his contemporary followers the Self is just an illusion.

However, regardless of these ontological disparities, those who believe in the existence of a Self convert to two simple theses about its nature. The first may be formulated as a conditional: if there is a Self, it is the active and spontaneous principle that organizes, at least in part, the intellectual and rational activity of the cognitive system. The second is the metaphysical relation that everyone has to himself, namely identity. Thus, my assumption is that learning bases the entire process of recruitment on this special metaphysical relation everyone has to himself: identity. Because I *am* the cognitive system whose brain states are recruited to represent myself, there is no need for me to identify myself as the individual referred to by the representation. There is no question for me when I think that I am referring to myself rather than to someone else.

There is enough to reconsider Baker's objection to naturalism:

Suppose that Jones is being tested on her ability to read PET scans, and that she is reading a contemporaneous PET scan of her own brain. Although she knows that she is reading a PET scan of an alert subject named "Jones," she does not realize that that Jones is herself*. (She thinks that the Jones whose brain she is watching is in the next room.) Now suppose that the telephone rings in the next room, where she thinks that the subject Jones is located. At *t*, she points to a lit-up portion of the brain on the screen and says to the tester, "Now Jones is hearing the phone." In so saying, Jones is expressing her thought that now Jones is hearing the phone. At the same time, with no conscious inference, Jones thinks to herself, "Jones is having the thought that Jones is hearing the phone." (1998, p. 28)

Being told that she is reading a contemporaneous PET scan of someone else, when Jones hears the telephone ringing and points to a lit-up portion of the brain on the screen and says to the tester, she thinks:

23. Now Jones is hearing the telephone.

However, just observing the PET scan she cannot realize that:

24. I am the one hearing the telephone.

My reply is pretty easy: Jones is unable to infer the truth of 24 from the truth of 23 because the channel of information about the truth of 23 is not internal, that is, in 23 she represents herself but not based on the metaphysical relation of identity that she has to herself.

From this, it follows that Jones has more than one concept of herself. We have two concepts with the same content. Jones's failure to recognize the person hearing the telephone ringing as herself is essentially like the

failure of early astronomers to recognize Hesperus as Phosphorus. One is the genuine self-concept linguistic expressed by the employment of tokens of the first-person pronoun in sentences like 24. The other (in Baker's case) is the concept expressed linguistic by her proper name "Jones" in sentence like 23. Again, the differences between those concepts of herself are not accounted for in linguistic or semantic terms, but rather in terms of the cognitive architecture. The genuine self-concept emerges from the selfless recruitment of the available self-information provided by internal channel to represent herself as such, based on the fundamental metaphysical identity relation that Jones has to herself. In contrast, the second concept of herself emerges from the recruitment of the self-information provided by her observation of the PET scan to represent herself as Jones.

The Content of Self-consciousness

As it is well known, Lewis proposes that we abandon the traditional theory of contents, taking them to be properties instead of propositions. Contents would be entities that are true or false, given a full characterization of a way for the world to be, only relative in addition to a subject and a time. Alternatively, the contents of propositional attitudes are, or at least select, not just classes or worlds, but rather classes of centered worlds: worlds together with a designated subject and a time. In coming to believe what he would express by accepting the sentence

25. I am making a mess,

Perry locates himself among all subjects making a mess at a given time and in a given world. However, Perry formulates the following objection to Lewis's relativistic or centered world account:

I believed that certain proposition, that I am making a mess was true—true for me. So belief that this proposition was true for me then does not differentiate me from some other shopper, who believes that I am making a mess, was true for John Perry. So this belief cannot be what explains my stopping and searching my cart for the torn sack. Once we have adopted these new-fangled propositions, which are only true at times for persons, we have to admit also that we believe them as true for persons at times, and not absolutely. And then our problem returns. (1979, p. 44)

However, this is a misunderstanding of Lewis's proposal that the content of Perry's epiphanic belief expressed by sentence 21 is *the property of making a mess*, which, in judging it, he self-ascribes (at the time of believing) the way ordinary propositions are supposed to be ascribed to the actual world when they are believed. The other shopper does not believe (i.e., self-ascribe) this property; she instead believes a

traditional *de re* proposition: that the property making a mess applies to Perry.

Be that as it may, the best way to capture the emergence of self-consciousness on a natural basis is Lewis's opposition to Perry's semantic account. What I want to suggest is the following. Before the emergence of self-consciousness, we must assume that the Self in this case is not a constituent of the representational content, but rather of the circumstance of evaluation of the property. For example:

26. The property of making a mess.

The content of sentence 26 is true only in the case that the property of making a mess is true in the world centered on John Perry at the time he is making a mess. If we assume the theory of possible worlds as a propositional model, then the representational content of sentence 26 (which does not yet represent itself in the content) will be the set of the centered worlds in which such a property is instantiated in the world centered on Perry and at the time he is making a mess. That content is true only in cases when the actual world is a member of the set of centered worlds.

Now, things change dramatically when brain states are recruited through learning with the function of indicating the very subject entertaining the perceptual thought expressed by sentence 26. From now on, Lewis's semantics are no longer useful. Rather than a *centered property*, what we now have is a complete *de se* proposition 25.

Again, if we assume the theory of possible worlds as a propositional model, then the representational content of sentence 25 will represent the set of possible worlds in which Perry is making a mess. That content is true only in cases in which the actual world is a member of the set of those possible worlds.

The remaining question is why we need *de se* thoughts. Let us assume the story of Crusoe. The brain states of Robinson Crusoe that convey self-specifying information about his body, about his perceptual field, do not have the function of indicating him just because *he is an argument-role that never changes*. The analogy here is with time zones. Before the Europeans were in America, they never had to worry about time zones since these were parameters that never changed. Things change when we are communicating with people from different continents.

Likewise, when Crusoe meets Friday, Crusoe's situation changes dramatically. Now Crusoe is no longer the only parameter. He needs to communicate his experiences, thoughts, etc. to Friday. And for this he needs to refer to himself as the subject of such experiences and thoughts. The states of Crusoe's brain that convey self-specifying information are

now recruited with the function of presenting Crusoe to Friday on the basis of the relationship of identity he has with himself. It is only from that moment that we can speak of knowing self-reference.

Notes:

1. The German term refers to the Latin “radix” meaning "root", so "root capacity" would be a reasonable translation.
2. Tugendhat (1979) is the main victim of the illusion of the linguistic turn as a solution to the puzzle of self-consciousness. But Wittgenstein (1958) and Shoemaker (1968) had already planted the seeds. Indeed, a similar puzzle was also recently formulated in linguistic terms. To master the token-reflexive rule of employment of the first-person pronoun the speaker must be already conscious of herself as the producer of the relevant token of that pronoun in order to realize that by employing it she refers to herself. See Bermúdez (1998).
3. As we will see, Baker’s challenge can be understood in quite a different sense. Only one of them is useful for my account.
4. For different reasons, Tugendhat reached the same formula: “His immediate knowledge of his conscious states is expressed by statements of the form: “I know that I” and any predicate may follow that expresses the possession of a state of consciousness. Thus, for example, “I know that I am bored”, “I know that I no longer intend to attend this lecture and so on.” (1979, p. 14).

References

- Baker, L. R. (1998). "The First-Person Perspective: A Test for Naturalism" in *American Philosophical Quarterly*, 35(4), 327–348.
- Baker, L. R. (2012). "From Consciousness to Self-consciousness" in *Grazer Philosophische Studien* 84, 19–38.
- Bermúdez, J. L. (1998). *The Paradox of Self-Consciousness*. Cambridge, USA: MIT Press.
- Block, N. (1995). "On confusion about a function of consciousness" in *Brain and Behavioral Sciences* 18 (2): 227-247.
- Burge, T. (1979) "Individualism and the Mental" in *Midwest Studies in Philosophy* 4: 73-122.
- Castañeda, H. N. (1966). "He: A Study in the Logic of Self-Consciousness" in *Ratio*, 8(2), 130–157.
- Chalmers, D. J. (1996). *The Conscious Mind: Toward a Fundamental Theory*. (Oxford, USA: Oxford University Press).
- Cramer, K. (1974). *Erlebnis. Thesen zu Hegels Theorie des Selbstbewußtseins mit Rücksicht auf die Aporien eines Grundbegriffs nachhegelscher Philosophie*. In H. G. Gadamer (ed.), *Stuttgarter Hegel-Tage*. Stuttgart, Germany: Reclam
- Evans, G. (1982). *The Varieties of Reference*. Oxford, USA: Oxford University Press.

- Frank M. (2002). *Self-consciousness and self-knowledge: On some difficulties with the reduction of subjectivity*. *Constellations*, 9(3), 390-408.
- Freud, S. (1999). Die Zerlegung der psychischen Persönlichkeit – Neue Folge der Vorlesung zur Einführung in die Psychoanalyse, in *Gesammelte Werke – Chronologisch geordnet*, Frankfurt am Main, Germany: Fischer Verlag.
- Dretske, F. (1995). *Naturalizing the Mind*. Cambridge, USA, MA: MIT Press/Bradford Books.
- Henrich, D. (1967). Fichtes ursprüngliche Einsicht. In D. Henrich und H. Wagner (eds.). *Subjektivität und Metaphysik*. Festschrift für Wolfgang Cramer. Frankfurt, Germany: Frankfurt a. M.: 188-233.
- Hume, D. (1967). *A treatise of human nature*. London, UK: Oxford University Press.
- Kant, I. (1956). *Kritik der reinen Vernunft*. Hamburg, Germany: Felix Meiner Verlag.
- Lewis, D. (1979). "Attitudes De Dicto and De Se". *Philosophical Review* 88, 513–543.
- Metzinger, T. (2003). *Being No One: The Self-Model Theory of Subjectivity*, Cambridge, USA, MA: MIT Press.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review* 83: 435-50.
- Perry, J. (1979). "The essential indexical" *Nous*, 13, 3-21.
- Pothast, U. (1971). *Über einige Fragen der Selbstbeziehung*. Frankfurt, Germany: Frankfurt a. M.
- Shoemaker, S. (1968) "Self-Reference and Self-Awareness" *Journal of Philosophy*, 65, 555–567.
- Tugendhat, E. (1979). *Selbstbewusstsein und Selbstbestimmung. Sprachanalytische Interpretationen*. Frankfurt, Germany: Frankfurt a. M.
- Zahavi, D. 2006. *Thinking about (Self-) Consciousness: Phenomenological Perspectives*. In U. Kriegel & K. Williford (eds.). *Self-Representational Approaches to Consciousness*, 273–295. Cambridge, USA, MA: The MIT Press.
- Wittgenstein, L. (1958). *The Blue and Brown Books: Preliminary Studies for the Philosophical Investigations*. Oxford, UK: Blackwell.