Metaphysical Self-identity without Epistemic Self-identification – A Cognitivist Solution to the Puzzle of Self-consciousness

Roberto Horácio de Sá Pereira

Abstract

This paper presents a new cognitivist account for the old puzzle of self-consciousness or knowing self-reference. Knowing self-reference does not rely on reflection on some putative pre-existent pre-reflexive self-consciousness nor is it the result of a process of identification of oneself as the employer of the relevant token of "I" according to the token-reflexive rule of the first-person pronoun. Rather, it relies on the architecture of the cognitive system. By exploiting the acquaintance-relation that every brain has to one's own body, a "mental file" is opened in a subliminal way to house information about oneself gained through the appropriate proprioceptors. Knowing self-reference based on self-files dispenses with self-identification because the opening of the file in the same brain/body and it is a not a deed of the cognitive system, but rather something that takes place sub-personally.

Prolegomena

Since Fichte, self-consciousness (knowing self-reference) has been quite embarrassing to the continental tradition. In a nutshell, to knowingly self-refer, I have to identify myself as the subject performing the very act of self-reference in the first place. But that launches a regress or presupposes self-consciousness in turn rather than accounting for it. To avoid the puzzle, the phenomenological tradition has postulated the existence of an intransitive form of self-consciousness: traditional self-consciousness results from reflection on some putative omnipresent pre-reflexive form of self-consciousness. In contrast, the members of the so-called "Heidelberg school" despaired of ever solving it. Henrich describes the phenomenon of self-consciousness as an "enigma" (1966, p. 65), and characterizes the philosophical attempt to explain the phenomenon as completely "helpless." Cramer remarks that the phenomenon of self-consciousness confronts us with "an incontestable state of affairs" whose explanation leads to difficulties that "appear close to insurmountable" (1974,

54). In a similar vein, Pothast (1971) describes the main difficulty of reflexive self-reference as "insoluble."

In the late seventies, Tugendhat (1979) claimed that the problem lurking behind the theory of reflection is what he, inspired by Heidegger (1989), called the subject-object model, that is, the underlying assumption that one becomes conscious of something insofar as one represents <*Vorstellen*> and identifies it as an object, i.e., as something placed against one's act of reflection. Tugendhat's diagnosis is correct in that the puzzle relies on a misunderstanding: no self-identification is required. Yet, his positive account is unconvincing. Tugendhat believed that he could solve or dissolve the puzzle by just replacing the subject-object model with his linguistic turn ("language-analytic approach"). However, the traditional puzzle can also be easily formulated in linguistic terms (see Bermúdez 1998). That said, it must be added that all discussed proposals just presuppose and do not explain. That is my aim in this paper: to account for self-consciousness in way that does not presupposes it and avoids any regress.

This paper presents a new cognitivist account for the old puzzle of knowing self-reference. Knowing self-reference does not rely on reflection on some putative pre-reflexive self-consciousness nor is it the result of a process of identification of oneself as the employer of the relevant token of "I" according to the token-reflexive rule of the first-person pronoun. Rather, knowing self-reference relies on the architecture of the cognitive system itself. By exploiting the acquaintance-relation that every brain has to one's own body, a "mental file" is opened in the brain to house information about oneself gained through proprioceptors. Given this, knowing self-reference based on self-files dispenses with self-identification because the opening of the file in the brain is a not a deed of the cognitive system, but rather something that takes place sub-personally.

This paper is structured as follows. Besides this brief introduction, the formulation of the puzzle extends to the next section. As my aim is to criticize the linguistic solution, the following two sections are devoted to presenting and criticizing Tugendhat's "language-analytic" approach. I have chosen Tugendhat's approach because it is by far the most detailed and clear in the literature. In the fourth section, I rule out Bermúdez's and the traditional phenomenological accounts. In the last section I present my own proposal.

I. The Puzzle

Self-consciousness is the ability to knowingly self-refer as opposed to different forms of self-reference without knowledge. As such, it seems to require the knowledge of oneself (object) as the very subject making the self-reference. The idea that consciousness requires a meta-representation belongs to a very old tradition, the theory of reflection, namely a turning-back-upon-oneself, which dates back to Ockham and continues through Descartes, Locke, Leibniz, until Kant and his followers. E.g., I am consciously representing this flower in the garden when I am able to represent (through a thought or a further perception) this sensible representation (meta-representation) as my representation. The problem seems to be that in order to knowingly self-refer I have to know, *previous to the act of self-reference*, that I am the subject (and the object) performing the very act.

Fichte was the first to recognize a puzzle in the account of self-consciousness by means of the theory of reflection (Henrich, 1966, p. 14). Knowing self-reference requires the knowledge that it is the object of the reflection that is at the same time the very reflecting subject, that is, the one who is performing the act of reflection. In Henrich's words:

> 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, 11, emphasis added)

The putative puzzle can be more clearly reconstructed in the form of a classic dilemma. The first arm of the dilemma is the infinite regress. The question is: how do I know that I myself am the object (of my own reflection)? The answer is: by knowing that I am the one who carried out the act of reflection in the first place. But the same question is raised all over again: how do I know that I am the subject who carried out the first-order reflection? (How do I know that I am the producer of the relevant token of the first-person pronoun?) For that, I need to perform another second-order reflection in order to identify myself as the subject who carried out the first-order reflection and so on *ad infinitum*.

The other arm of the dilemma is the vicious circle. If I want to avoid the undesirable vicious infinite regress, I have to assume that I somehow know in advance that I am the same subject who is the object of reflection by knowing, and at the same time, *and by the same token of the first-person pronoun*, that I am the subject that has carried out the act of first-order reflection (without the

need to carry out a second-order reflection). However, if I already know that I am the subject carrying out the act of reflection (the one producing a token of the first-person pronoun), as Fichte says, the knowing self-reference is not explained by the theory of reflection but rather it is presupposed.

Fichte's own solution to this problem is unclear: "self-positing." In fact, according to Henrich, Fichte never explained his metaphor of positing and self-positing (1966, 18). The formula "the 'I' posits itself" can only negatively characterize Fichte's own rejection of the theory of reflection. However, according to the Heidelberg school, the idea of "self-positing" sounds incomprehensible. Pothast wonders: "how could someone perform that very act of positing if it does not yet exist in the first place" (1971, 71)?

Thus, the Heidelberg theory of consciousness emerged in 1971 as an attempt to resume Fichte's original insight. Negatively, it can be characterized by the rejection of both the old theory of reflection and Fichte's claim that self-consciousness is a sort of intellectual intuition of the sheer activity of apperception. Positively, it can be seen as the resumption of Fichte's original insight that self-consciousness must be based on a non-propositional knowledge of oneself, which Henrich calls self-acquaintance.

The core of the old Heidelberg theory of consciousness can be represented in three main theses. (1) Reflexive knowing self-reference cannot be accounted for in the terms suggested by the theory of reflection without circularity. (2) To break the circle, self-consciousness must be accounted for on the basis of an original form of self-acquaintance within consciousness. (3) This original form of self-acquaintance is neither an activity nor a relation between a subject and her object. However, it remains a mystery what (2) and (3) mean exactly and how they are supposed to solve the old puzzle.

II. "Language-analytic Turn"

Inspired by Heidegger (1989), Tugendhat claims that Fichte's puzzle arises only because self-consciousness is misconceived in the traditional terms of the subject-object model of consciousness. To be sure, I become aware of this computer by means of some intentional act of representing it as an object. Still, I do not become conscious of myself by means of some intentional act of representing me as an object. Tugendhat summarizes his criticism of the traditional view of reflexive self-reference in the following terms:

> The problem with the theory of reflection that Henrich identifies (…) rest

> on the assumption that we are analyzing something whose essence consists in the identity of knowing and what is known. For someone who does not acknowledge that the phenomenon of self-consciousness has or presupposes this structure, the difficulty does not exist. The difficulty, which is in fact insoluble, is only an outcome of the absurdity of the basic approach. (1979, p. 64)

In Tugendhat's view, the problem of the theory of reflection traces back to the subject-object framework. The puzzle only emerges because self-consciousness is misconceived as the alleged relation of identification between "I" *as the representing subject*, and me *as the represented object*, which results from a self-representation. In other words, the background assumption is that one becomes conscious of oneself by self-identifying oneself as the object of one's own intentional act of representing. In this regard, Tugendhat is quite right. We will come back to this point in the last section.

Yet, Tugendhat's solution involves a methodological re-orientation toward language:

> One asks oneself whether this problem disappears – or at least can be solved in any case – under the language-analytical view of epistemic self-consciousness, understood as that view that proceeds from the assumption that epistemic self-consciousness manifests itself in language, instead of relying on inner awareness. (1979, 54)

Tugendhat's "language-analytical approach" is characterized by two closely connected tenets. The first – the negative one – is his rejection of the subject-object framework; the second – the positive one – is his "language-analytical" reduction of the reflexive self-reference phenomenon to the mode of employment of psychological I-sentences in which one takes a self-ascription of a mental predicate "@." So, the understanding of reflexive or knowing self-reference relies on the understanding of the mode of use of the first-person pronoun and on the mode of employing mental predicates.

As the ultimate reference point, the first-person pronoun doesn't identify or pick myself out as one among other individuals in some domain. The lurking question is why. Wittgenstein and Anscombe notwithstanding, Tugendhat holds that the first-person pronoun does refer to my person as someone identifi*able* from the third-person perspective. Given this, any sentence "I @" does express a genuine proposition rather than a mere avowal <Äusserung>. On this basis, Tugendhat states what he calls the semantic *principle of veridical symmetry* between first-person and third-person psychological sentences:

The sentence "I @" is true, if uttered by me, iff the sentence "He @" is true if uttered by someone else who by "he" means me <mich meint> (Tugendhat, 1979, p. 88).

According to Tugendhat, what ensures the veridical symmetry is the reasonable assumption that the indexicals "I" and "he" involved co-refer. When someone self-refers by means of the first-person pronoun and when someone else (or the person himself) refers to that person by means of the third-person pronoun, *one and the same proposition is being expressed*:

1. He (Ernst) feels pain,

And what Ernst says or thinks

2. I (Ernst) feel pain.

Yet, the simple co-reference of the indexicals involved is necessary but certainly not enough for the veridical symmetry. Tugendhat overlooks a key assumption. It is also necessary that we take a coarse-grained Russellian proposition as the appropriate model for the content of 1 and 2, in this case a sequence consisting of <Ernst; Pain>.

Now, although one and the same Russellian proposition is being expressed by 1 and 2, <Ernst; Pain>, it is only by thinking 2 that Ernst knows without identification that he is self-referring. In opposition to 2, Ernst's or someone else's knowledge of the truth of 1 is based either on the observation of Ernst's behavior (when the thinker of 2 is someone else) or, in some cases, on inferences. In this way, the principle of veridical symmetry requires that the content expressed by 1 and 2 is modeled by Russellian propositions.

III. The Puzzle Returns

But if the immediate knowledge of oneself as the owner of mental states is negatively described as not based on observations, inferences, and on alleged inner perception, Tugendhat owes us a positive explanation of it. Following Wittgenstein and Shoemaker, Tugendhat holds that psychological first-person sentences are immune to a peculiar error of reference when employed in conformance to the rule. So, if Ernst knows the rule of employment of the first-person pronoun (according to which that pronoun refers to whoever employs a token of it), by employing a token of it, Ernst couldn't possibly fail to recognize that he is referring to himself whenever he thinks 2.

Yet, Tugendhat's equation of immediate epistemic self-consciousness and

the employment of psychological I-sentences in conformity to its rule raises several questions. First, what guarantees the immediate self-knowledge of the content expressed by 2 is certainly not the Russellian proposition consisting of the sequence <Ernst; Pain>, but rather the mastering of the token-reflexive rule of the employment of the first-personal pronoun. Given this, the appropriate model for capturing the immediate self-knowledge expressed by the content of 2 is some Fregean proposition consisting of the peculiar mode of presentation of Ernst's expressed rule of employment of "I," roughly:

3. The individual employing a token of 2 (Ernst) is in pain.

The meaningful employment of 3 relies on what Bermúdez has called the token-reflexive rule of the employment of the first-person pronoun:

4. If a person employs a token of 'I', then he refers to himself in virtue of being the producer of that token. (Bermúdez 1998, p. 15)

Let us assume just for the sake of argument that Ernst is on the street when he calls his mom to complain about his pain. Since his mother is not at home, the answering machine is automatically activated and Ernst utters sentence 2, recording it on the answering machine. Time goes by and Ernst's pain is over and he forgets about it. He then returns home and checks the messages on the answering machine. However, when he listens to the messages from the answering machine, Ernst does not recognize his own voice. What conclusion can we draw from this simple case? First, Ernst must assent to the content of sentence 3 (what he listens to when he hears the messages from the answering machine), provided he only masters the semantic of the rule of the employment of the first-person pronoun 4. Yet, at the same time, he can deny the content of 2, modeled as a Russellian proposition, even though the contents of 2 and 3 are veritatively symmetrical. Even worse, as Ernst does not recognize his own voice on the answering machine, even when he assents to 3, he is not knowingly or reflexively self-referring. Reflexive self-reference requires the employer of the first-personal pronoun to have knowledge of his own identity.

The problem is: as Ernst was not born knowing of rule 4, how did he come to master it if he was not already self-conscious in the first place? Interestingly, Henrich gives this reply to Tugendhat's criticism by claiming: "if we understand the word ("I") as an indexical word, the problem is eliminated" (1970, 49); that is, the problem is presupposed rather than solved and we are back at Fichte's puzzle. But Tugendhat never took Henrich's reply seriously because he never

understood that the puzzle is cognitive rather than linguistic.

The following conclusions are imposing. To be sure, Tugendhat is right when he claims that self-consciousness cannot rely on the traditional subject-object model. Nobody becomes self-conscious by *identifying* himself as the object of his own intentional act. Moreover, a Fichtean intellectual self-intuition is out of the question here. Nevertheless, to appeal to the token-reflexive rule 4 *presupposes* rather than solves the problem because in the token-reflexive rule 4 self-identity is presupposed rather than explained. We are back at Fichte's puzzle: the employment of tokens of the "I" presupposes reflexive or knowing self-reference rather than explaining it.

Tugendhat's greatest mistake was to assume the original puzzle was linguistic rather than cognitive. Indeed, there is a revival of the theory of reflection in analytical philosophy under the label of higher-order theories of consciousness (see Rosenthal 1986, 1993, 2004, 2005, 2011, 2018; Carruthers 1989, 1996, 1999, 2000; Lycan 1987, 1996, 2001a, 2001b, 2004, and Gennaro 1996, 2004, 2005, 2006, 2012, 2015).

The imposing conclusion is that the "linguistic turn" cannot solve the problem. Rather, it presupposes it as solved by assuming that by employing the first-personal pronoun the subject self-refers knowingly without the need for self-identification. Tugendhat's approach is a paradigm of the Wittgensteinian way of solving philosophical problems.

IV. Roads not taken

The common idea of the Heidelberg school and of the phenomenological tradition is that before reflexive or knowing self-reference takes place by means of mastering the token-reflective rule of the first-person pronoun the individual is "already somehow acquainted with himself." The idea might sound right. However, it is nothing more than a metaphor! The question is: how to provide a positive meaning for this metaphorical claim? In this section I mention roads I believe that we should not take.

1. In his famous book, Bermúdez (1998) believes that the only solution to this traditional puzzle is the postulation of primitive nonconceptual forms of self-consciousness. To be sure, I believe that a primitive nonlinguistic form of self-consciousness is necessary, but I do not see why this nonlinguistic self-consciousness must at the same time be nonconceptual. We are back at the linguistic turn! Bermúdez rightly rejects what he calls *The Conceptual Require-*

ment Principle, making room for the possibility of nonconceptual contents:

> *The Conceptual Requirement Principle:* The range of contents that one may attribute to a creature is directly determined by the concepts that the creature possesses. (1998, 41)

However, he is still closed to the linguistic dogma when he accepts the priority principle:

> *The Priority Principle:* Conceptual abilities are constitutively linked with linguistic abilities in such a way that conceptual abilities cannot be possessed by nonlinguistic creatures. (1998, 42)

On the phylogenetic scale, genuinely perceptual systems appear in animal species well before belief and propositional attitudes appear. Bees, frogs, pigeons, goldfish, and octopi are, I assume, good examples. Although they lack propositional attitudes, they have visual perceptual systems. The perceptual systems of some of these animals have been thoroughly studied. Scientific explanations of the discriminations, computations, and informational functions of the perceptual systems of lower animals commonly individuate the representational content of visual states partly in terms of properties and relations in the animals' environment, properties and relations to which the animals bear causal relations – both in sensory reception and in activity. In fact, the best explanations of some of these low-level representational systems attribute perceptions of physical objects in space, and of rudimentary spatial features of and among such objects. For example, computations in the visual system of bees that bear on locating a hive operate on parameters that represent spatial positions and objects in those positions.

Yet, there are overwhelming data supporting the assumption that primates and other higher mammals have propositional attitudes-beliefs, conceptualized wants, and intentions – as well as perceptions. Having beliefs requires having a capacity for inference-for truth- and reason-preserving propositional transitions among propositional attitudes, transitions that are attributable, as activity, to the whole animal. Simple logical, inductive, and means-end inferences are present in the mental activity of higher non-human animals.

Moreover, I also assume that primates and other higher mammals that are known to have propositional attitudes-beliefs also have self-notions. A prey cannot think that a predator is coming towards it unless it has a self-notion. Of course, the possession of a self-notion does not mean that the creature

knowingly self-refers because without communication there is no need for self-reference in the first place. Thus, I do not see any compelling reason to assume that pre-linguistic infants that are about to learn token-reflexive rule 4 do not possess a self-notion, self-concept, or self-file. But what is a self-file? I will come back to this notion in detail in the last section of this paper.

For the time being, I wonder why this self-notion is not a nonconceptual self-reference. For one thing, the self-notion meets Evan's generality constraint for concept attribution: self-notions can be combined and recombined with quite different general properties and relations (Evans 1982). For example, the same prey A that thinks that a predator B is coming in its direction, may also think, now as a predator, that its prey C is within A's reach. In other words, A can think that B is coming towards A, but also think that A is coming towards C. If this is conceivable, A must possess a primitive pre-linguistic self-concept. The conclusion is the same as before: the "linguistic turn" has got everything wrong again!

2. As a way out of the dilemma, the phenomenologist postulates a pre-reflexive, intransitive form of access to oneself. In such primary self-disclosure, one doesn't take oneself *as an object* either of one's own inner perceptions or of one's own thoughts. According to Sartre, for example, it is only the necessity of syntax that compels us to say that we are aware *of* our experiences or *of* ourselves. The basic claim is that one's experiences and thoughts rely upon a peripheral awareness of oneself. When he focuses his attention on some cigarettes (Sartre's example), at the same time that he becomes transitively aware that they are twelve in number, he is also pre-reflexively aware that he is counting them. There is no infinite regress since, according to Sartre, "there is an immediate, noncognitive relation of the self to itself" (1956, 12).

Nonetheless, even if the postulation of a pre-reflexive or intransitive form of self-consciousness avoids the traditional puzzle because there is no need of identification, that is no solution to our problem insofar as the reflexive self-reference (i.e., the fully-fledged self-consciousness) is still understood in all phenomenological traditions as the result of a self-identification (the subject-object model). Sartre is quite explicit on this point: "[Reflection] is an operation of the second degree...performed by an [act of] consciousness *directed upon consciousness*, *a consciousness which takes consciousness as an object*" (Sartre, 1957, p. 44, emphasis added). So, if Sartre is pre-reflexively aware that he is counting (without taking himself as an object) while he sees some cigarettes, he could only become reflexively conscious of himself by counting when he takes and identifies himself as the object of a second-order consciousness. We are back at the regress.

V. The Cognitivist Proposal

The rational core of Tugendhat's criticism is his rejection of the traditional subject-object model of self-consciousness. Yet, Tugendhat overlooks that if self-consciousness does not result from a self-i*dentification*, that cries out for an explanation. To claim that uses of the first-person pronoun directly refer without the need of identification is just a restatement of the very problem/puzzle: why by producing a token of the first-person pronoun do I knowingly self-refer without the need of self-identification?

My first assumption is that we are primitively self-acquainted in proprioception, bodily sensations, feeling, and kinesthesis since the moment we are born. But what is self-acquaintance? For Russell acquaintance means several things, but what concerns me here is the thing-knowledge that dispenses with knowledge of truths. So, there is no need to propositionally identify myself in knowing self-reference. There is no need whatsoever to identify myself as the subject of proprioception, bodily sensations, feeling, and kinesthesis. But why is this so? What else can I say about self-acquaintance?

When I talk about self-acquaintance, the first thing that I have in mind is the architecture of the cognitive system. The brain is connected through proprioceptors to the body and bodily limbs and organs in such a way that whenever I feel pain, whenever I am standing, whenever I am running, etc. the information-flow is carried through proprioceptors to the cortex. In this way an integrated body schema starts to form. Even though we may be conscious, for example, that our legs are crossed, or that we are standing, the flow of information is processed below the threshold of consciousness. Moreover, in great part, the body schema is not even conscious. Think about our sense of bodily equilibrium. We first become conscious of it when we lose it. Thus, by self-acquaintance I mean the non-conscious processing of information-flow from the proprioceptors to the brain that dispenses with self-identification because it is based on the fundamental metaphysical identity between the brain and the body in the first place.

Now, by exploiting those proprioceptive channels between the brain and the body's limbs and organs, a mental file is opened inside the cortex to house the great amount of proprioceptive, kinesthetic, and bodily information. Let me call this *the self-file*. Again, that is something that nobody does as if there were a homunculus inside the brain, but rather something that happens below the threshold of consciousness. Indeed, before the self-file, the more or less integrated cognitive system is not even a subject. The person or subject is what emerges as the result of opening the file. Only at this very moment is the body schema or bodily self-representation accomplished.

In the beginning, that file is temporary like *a buffer* in a computer: it lasts just a few hours or even less. In this case we may talk about a nonconceptual form of self-awareness: the buffer as a temporary file cannot be recombined with representations of properties of which the cognitive system has a concept (Evans's generality constraint, see Evans 1982). However, it is easy to imagine that what was once a buffer turns little by little into a real self-file over time, housing all of sorts of information about oneself. At this moment, much before the acquisition of language, and mastering of the first-person pronoun, we can appropriately talk about self-consciousness or knowing self-reference.

It is reasonable to assume that, initially, the file only houses information obtained through proprioceptors (that is what Tugendhat calls epistemic immediate self-consciousness). Afterwards, the self-file turns into a really singular self-concept since it can be recombined with any sort of information about the person's properties of which the person has a concept (Evans's generality constraint). Over time, the file houses information obtained in the third personal way, e.g. about the age, the height etc. of the person in question (that is what Tugendhat calls epistemic immediate self-consciousness).

Now, let me return to our puzzle of self-consciousness. The problem only emerges because to knowingly self-refer I have to *identify* myself as the subject performing the act of self-reference. But that seems to require the previous knowledge – *previous to the act of self-reference* – that I am the subject performing the very act.

Nonetheless, no such self-identification or previous knowledge is required. For one thing, as we saw, the flow of information through the proprioceptors that is in part responsible for the body's self-schema is processed in a subliminal way, below the threshold of consciousness. For another, behind the opening of the mental file about oneself in the brain, housing all sorts of information about oneself, there is no homunculus. The opening is not a deed of the system, but rather something that happens to it.

References

Bermúdez, J.L. 1998. *The puzzle of self-consciousness*. Cambridge: MIT Press.
Carruthers, P. 1989. "Brute experience." *Journal of Philosophy*, 86: 258–269.
——, 1996. *Language, Thought and Consciousness*. Cambridge: Cambridge University Press.
——, 1999. "Sympathy and subjectivity." *Australasian Journal of Philosophy*. 77: 465–482.
——, 2000. *Phenomenal Consciousness: a naturalistic theory.* Cambridge: Cambridge

University Press.
Cramer, K. 1974. Erlebnis, in: *H. Gadamer, Stuttgarter Hegel Tage*. Bonn.
Fichte, J.G., 1794. *Grundlage der gesamten Wissenschaftslehre*. Jena und Leipzig.
Gennaro, R. 1996. *Consciousness and Self-Consciousness.* Amsterdam: John Benjamins.
——, 2004. "Higher-order thoughts, animal consciousness, and misrepresentation." In R. Gennaro (ed.) 2004, pp. 45–66.
——, (ed.) 2004. *Higher-Order Theories of Consciousness*. Philadelphia: John Benjamins.
——, 2005. "The HOT theory of consciousness: between a rock and a hard place." *Journal of Consciousness Studies*, 12: 3–21.
——, 2006. "Between pure self-referentialism and the (extrinsic) HOT theory of consciousness." In Kriegel and Williford (ed.) 2006.
——, 2012. *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. Cambridge, MA: MIT press.
——, (ed.) 2015. *Disturbed Consciousness: New Essays on Psychopathology and Theories of Consciousness*. Cambridge, MA: MIT press.
Heidegger, M. 1989, *Die Grundprobleme der Phänomenologie*, Frankfurt a. Main.
Henrich, D. 1966. *Fichtes ursprüngliches Einsicht*. Frankfurt a. M.
— 1970: "Self-consciousness, a critical introduction to a theory." *Man and World* 4 (1): 3-28.
— 2007: "Selbstsein und Bewusstsein." http://www.jp.philo.at/texte/HenrichD1.pdf.
Husserl, E. 1973. *Zur Phänomenologie der Intersubjektivität III, Husserliana XV*. Den Haag Martinus Nijhoff.
Lycan, W. 1987. *Consciousness.* Cambridge, MA: MIT Press.
——, 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
——, 2001a. "Have we neglected phenomenal consciousness?" *Psyche*, 7. Available from the ASSC depository
——, 2001b. "A simple argument for a higher-order representation theory of consciousness." *Analysis*, 61: 3–4.
——, 2004. "The superiority of HOP to HOT." In R. Gennaro (ed.) 2004, pp. 93–114.
Pothast, U. 1971. *Über einige Fragen der Selbstbeziehung*. Frankfurt am Main: Vittorio Klostermann.
Rosenthal, D. 1986. "Two concepts of consciousness." *Philosophical Studies*, 49: 329–359.
——, 1993. "Thinking that one thinks." In Davies and Humphreys (eds) 1993.
——, 2004. "Varieties of higher-order theory." in R. Gennaro (ed.) 2004, pp. 17–44.
——, 2005. *Consciousness and Mind.* Oxford: Oxford University Press.
——, 2011. "Exaggerated reports: reply to Block." *Analysis*, 71: 431–437.
——, 2018. "Misrepresentation and mental appearance." *TransFormAcao*, 41: 49–74.
Sartre, J.P., 1956. *The Transcendence of the Ego*. Trans. Forrest Williams and Robert Kirkpatrick. New York: Hill and Wang
Shoemaker, S. 1996, the first-person perspective and other essays, New York, Cambridge University Press.
Tugendhat, E. 1979. *Selbstbewusstsein und Selbstbestimmung. Sprachanalytische Interpretationen*. English translation: P. Stern. *Self-Consciousness and Self-Determination.* Cambridge, MA: MIT Press, 1986.
Wittgenstein, 1958. *The Blue and Brown Books.* Oxford: Blackwell.

Roberto Horácio de Sá Pereira
Professor Titular / Distinguished Professor
Departamento de Filosofia / Department of Philosophy
Universidade Federal do Rio de Janeiro
Rio de Janeiro, RJ - Brazil