



Københavns Universitet



## First-Person and Minimal Self-Consciousness

Grünbaum, Thor

*Published in:*  
Consciousness and Subjectivity

*Publication date:*  
2012

*Document Version*  
Early version, also known as pre-print

*Citation for published version (APA):*  
Grünbaum, T. (2012). First-Person and Minimal Self-Consciousness. In S. Miguens, & G. Preyer (Eds.),  
Consciousness and Subjectivity (pp. 297-320). Heusenstamm: Ontos Verlag.

# FIRST PERSON AND MINIMAL SELF-CONSCIOUSNESS

*Thor Grünbaum*

## I. Introduction

Many philosophers accept that consciousness involves some form of very minimal self-consciousness. In its most general form the idea would be that phenomenal consciousness can present the world as being a certain way but does this by phenomenally presenting it to a subject. This relation to a subject is not an external description of a metaphysical fact. It is part of the phenomenal character of the experience. Here is how Ned Block describes this feature of phenomenal experience:

[he is arguing that there is a form of “self-givenness” more fundamental than full-blown self-cognition] But even if monkeys and dogs have no self-consciousness, no one should deny that they have P[henomenally]-conscious pains, or that there is something it is like for them to see their reflections in a mirror. P-conscious states often seem to have a “me-ishness” about them, the phenomenal state often represents the state as a state of me. (Block 2007b: p. 178)

And here is the phenomenon described by Tyler Burge:

Consider subjects which have phenomenal or intentional states, but which by their nature lack critical reason [...]. The very existence of perceptual states or sensations – even in the absence of propositional ability – requires a subject, an individual with subjectivity or consciousness. Perceptual systems of lower animals require a subject, and it is clear that those systems have some sort of non-conceptual egocentric sensitivity. [For all beings lacking a full first-person concept] these beings’ mental states require a subject, whose subjectivity is a necessary aspect of their sensations, perceptions, or propositional attitudes. [Conscious states and events] presuppose an individual subject with a subjective perspective. Egocentric sensitivities or concepts mark this perspective. (Burge 1998: p. 248)

And finally, here is how Dan Zahavi presents the idea:

In its most primitive and fundamental form, self-consciousness is taken to be a question of having first-personal access to one’s own consciousness; it is a question of the first-personal givenness or manifestation of experiential

life. [...] Insofar as there is something it is like for the subject to have the experience, the subject must in some way have access to and be acquainted with the experience. Moreover, although conscious experiences differ from one another [...] they also share certain features. One commonality is the quality of *mineness*, the fact that the experiences are characterized by first-personal givenness. That is, the experience is given (at least tacitly) as *my* experience, as an experience *I* am undergoing or living through. Given this outlook, it is natural to argue that self-awareness is of pertinence to an understanding of phenomenal consciousness. In fact, phenomenal consciousness must be interpreted precisely as entailing a minimal form or thin form of self-awareness. On this account, any experience that lacks self-awareness is nonconscious. (Zahavi 2005: pp. 15–16)

These philosophers endorse one version or another of what I will call the Theory of Minimal Self-Consciousness. In its strongest formulation, the theory says that phenomenal consciousness entails minimal self-consciousness, or what Block calls “me-ishness” and Burge “egocentric sensitivity”. Necessarily, if a person is phenomenally conscious (of something), then she is minimally self-conscious.<sup>1</sup>

It is important to notice that being minimally self-consciousness is not to be understood as a person’s being aware of herself. To be minimally self-conscious is not the same as being conscious of a self. Minimally self-conscious is not a new act of thinking or perceiving taking the self as its object. In Burge’s formulation it should be understood as a necessary aspect of any sensation, perception, and propositional attitude. It is thus not a mental state in itself, but an aspect or feature of any conscious state or event. Minimal self-consciousness does not require conceptual abilities, a first-person concept, or any form of attention. It only requires that the subject is phenomenally conscious. It simply concerns the fact that “the experiences are characterized by first-personal givenness” (Zahavi, *ibid.*).

A theory of minimal self-consciousness is accepted by a large number of philosophers. To many of them a theory of minimal self-consciousness seems to be self-evidently true. At least, only a small number of philosophers offer us explicit arguments in favour of a theory of minimal self-consciousness.<sup>2</sup> In the contemporary literature, we find arguments concerning how best to

1 Notice that Block endorses a weaker version according to which only some phenomenal states are characterized by “me-ishness”. Many philosophers with different and probably mutually inconsistent outlooks subscribe to the stronger claim that phenomenal consciousness entails a thin form of self-consciousness or egocentric sensitivity. Among them are Goldman 1970, 96, Frankfurt 1988: p. 162, Flanagan 1992: p. 194, Kriegel 2009, Siewert 1998, and Chalmers 1996.

2 For a recent exception, see Kriegel 2009.

account for minimal self-consciousness (roughly, should the account be two-level representationalist, one-level representationalist, or one-level non-representationalist), but rarely do we find any attempts to argue for the existence of minimal self-consciousness in the first place. One reason for this fact is, I think, that a theory of minimal self-consciousness looks like a near conceptual truth. If we understand minimal self-consciousness as first-personal givenness (the phenomenal character of me-ishness or egocentric sensitivity), then it becomes almost unintelligible what a denial of the claim that phenomenal consciousness entails minimal self-consciousness would amount to.

I do not think that the theory of minimal self-consciousness is a conceptual truth. Some philosophers explicitly deny the theory of minimal self-consciousness; others are implicitly committed to its falsity in virtue of their views on phenomenal consciousness. I will not assume that in their denial, these philosophers are uttering some form of nonsense. What we need is therefore some argument in favour of a theory of minimal self-consciousness. In this paper I will present one possible way of arguing for the theory of minimal self-consciousness, namely, by an argument by elimination.

*The argument:* Argument for minimal self-awareness:

- (1) Either a theory of consciousness is an anonymity theory or a theory of minimal self-awareness
- (2) If the anonymity theory is true, conscious experience is impersonal
- (3) If the theory of minimal self-awareness is true, then conscious experience entails minimal self-awareness
- (4) If a theory of consciousness cannot explain first-person self-reference, then the theory is false
- (5) An anonymity theory cannot explain first-person self-reference
- (6) Therefore, anonymity theory is false [from 4 & 5]
- (7) Therefore, the theory of minimal self-awareness is true [from 1 & 6]
- (8) Therefore, conscious experience entails minimal self-awareness [from 3 & 7]

The premises 1–3 are more or less definitional. Either a theory claims that consciousness entails minimal self-consciousness or it does not. If it does, then experience has as an intrinsic component some form of self-presentation. If it does not, then experience is impersonal. The central premises are 4 and 5. In the remainder of the paper I will sketch an argument in their favour. I will do this in the following way. First, I will describe certain facts about first-person thinking. I will argue that we should accept these facts. On any theory

these facts are something you must honour. If you deny them you have an explanatory problem. Second, I will argue that the anonymity theory cannot honour these facts. Consequently, the anonymity theory cannot explain the phenomenon of first-person thinking as this phenomenon is characterized by the basic facts.

## 2. Self-Conscious Thinking

Human agents think self-conscious thoughts. A thought can be said to be a self-conscious thought when a person consciously thinks about herself with an “I”-concept. Most of our thinking is probably directed at the world. But sometimes we think thoughts about ourselves. For example, in the case of self-identification: when I realize that I am the person in the picture or that I am the person making the sugar trail in the store. Or in the case of self-location and spatial cognition: when I compare my present visual relation to some landmark to the location of that landmark on my map and figure out which way to go next. Often the “I”-concept is more implicitly than explicitly present in one’s thinking but it is there nonetheless. For example, in the case of deliberation and decision-making: I might be deliberating how to get a piece of furniture through a narrow door opening, and my conclusion might be that it should be turned like this. But if I did not think that I must turn it like this if I want to get it through the door, then, intuitively, the content would have no motivational force for me.

There is a crucial difference between thinking a thought about oneself with an “I”-concept (thinks an “I”-thought) and thinking about oneself in other ways. There is thus an important difference between thinking “I’m bleeding from my hand” and thinking “TG is bleeding from his hand” (“this guy is bleeding from his hand” or “that hand is bleeding”) even though in all cases reference is being made to the same person’s hand. All the thoughts are made true by the same fact. Let us describe this distinction between thinking about oneself with a first-person concept and thinking about oneself in other ways with a distinction between first-person self-reference and third-person self-reference. I want to highlight three features characterizing our ability to make first-person self-reference in thinking.

First, a *semantic feature*. A person who thinks about herself with an “I”-concept and thereby engages in an act of first-person self-reference is guaranteed not only to refer to some object but to refer to herself. All judgements making use of the first-person concept share this important features. They are all guaranteed success in actually referring to the subject of thinking

(Castañeda 1968). It is impossible to be thinking an I-thought (i.e. to be making a self-conscious judgement) and not be referring to oneself. I can be radically deluded in my self-ascription (“I am the ruler of the universe” or “I am Napoleon”) but I am still ascribing the property to myself. If I am thinking that my hand is bleeding or that I am very intelligent, then there is no possible situation where I could end up actually referring to someone else. I could, of course, come to realize that I am in fact not the person who is bleeding (the bleeding hand belongs to someone else) or the one who is intelligent, and I should consequently stop thinking of myself as having the property in question. However, in so far as I am thinking of myself with an “I”-concept I am immune to reference failure.

It is important to distinguish between the fact that a judgement can be immune to error from misidentification (IEM) and the fact that it can be immune to error from reference failure. The two features of self-conscious judgements are independent of each other. The recent philosophical literature has had a tendency to focus on the first type of immunity (IEM). The reason for this has probably been that many philosophers have thought that immunity to error from misidentification has a special relation to the phenomenon of self-awareness (Shoemaker 1968, Cassam 1997). Others have, however, argued that IEM could not be the mark of self-awareness, since some judgements have IEM without being self-conscious judgements (e.g. thinking *this thing is red* when visually attending to a particular strawberry) and some judgements can be self-conscious without being IEM (e.g. I am bleeding from my hand) (Evans 1982, Howell 2007).

Second, *an epistemic feature*. A person cannot think an “I”-thought without understanding and knowing that she is thinking something about herself. When thinking an “I”-thought she is not only necessarily guaranteed success in referring to the right object, she also knows and understands that she is referring to herself and that she could not be referring to anyone else. By contrast, when referring to myself in the third-person mode (by a name, a perceptual demonstrative, or a definite description) there is always the danger that I am in fact referring to someone else or that I do not realize that I am in fact referring to myself (Castañeda 1967, Perry 1977). Thus even in the case where I am referring to myself by some definite description or name, it is possible that I could be thinking about myself without knowing and understanding that I am doing so. By contrast, if you can show that a person who is allegedly thinking an “I”-thought is in fact referring to someone else by her “I” or that she does not realize that she is in fact referring to herself, then you have simply shown that she was not thinking an “I”-thought at all.

Third, a *motivational feature*. There are important motivational differences between first-person self-reference and third-person self-reference in thinking. If I overhear people at work talking about some individual as being “obnoxious, self-pompous, and a miser”, it might not mean much to me and I might not care. But if it dawns on me that they are talking about me and that I am the individual who is “obnoxious, self-pompous, and a miser”, then it should matter a great deal to me and I would care. Or just imagine that you are looking at a TV-monitor, watching some person being approached from behind by a foul looking character with a raised knife. It would make a difference were you to realize that *that* is you on the screen. Ascribing some property to oneself by first-person self-reference thus relates the subject to the property in a special motivational way; in a way that matters to her emotional feeling, thinking, and practical reasoning. This was one of the facts Perry (1979) brought out with his famous example of himself in the supermarket suddenly realizing he was the person leaving behind a trail of sugar.

These three features of self-conscious thinking are relatively uncontroversial. They are intuitively correct and most philosophers accept them. I now want to add a fourth feature that I think is equally intuitive but which is far more controversial. The three features characterizing the first-personal self-reference of self-conscious thinking are (partly) explained by the way in which a conscious subject is presented to herself in conscious experience. That is, it is something about consciousness that (is part of what) explains our ability to think about ourselves in the first-person way.

This claim is controversial for the following reason. According to the one prevalent theory of mind, phenomenal consciousness is a non-functional and non-representational property. One way to understand this idea is by the example of blindsight. We can imagine a case of blindsight where the blindsight patient is behaviourally equal to a normally sighted person except for the fact that she claims that she is not able to see anything. The normally sighted person enjoys a conscious visual experience but the phenomenal consciousness is not part of what enables her rational thinking and acting (except for the subjective report). In other words, the visual phenomenal experience plays no cognitive role. Analogously, we could imagine a case of “self-blindsight” where the afflicted person and a normally functioning person are behaviourally equal to each other when thinking and acting in self-related ways. They can both think first-person thoughts. It is just that one does it non-consciously whereas the other one does it consciously. So, of course self-conscious thinking is conscious (by definition) but consciousness plays no enabling role for what is going on in self-conscious thinking.

If we accept that phenomenal consciousness is part of what enables a person to make first-person self-reference in thinking, then we have to accept that at least in the case of first-person thoughts phenomenal consciousness plays a cognitive and explanatory role. Non-conscious machines cannot think first-person thoughts even though we would want to attribute to them conceptual abilities. Just adding consciousness is not enough. Imagine a monoglot Englishman who is hypnotically induced to think “*Je suis un fou*”. This string of symbols would under their standard (French) interpretation be making self-reference. But in this case, the thinker is not making first-person self-reference. The intuition here is that the thinker is conscious and is tokening a sequence of symbols that under their standard interpretation would involve self-reference; yet, the thinker is not referring to herself in the first-person way because her consciousness and her thinking are not related in the right way. In a later section I will elaborate on this idea. For now, I will simply assume that it is something about phenomenal consciousness that is part of what explains our ability to self-refer in the first-person way. Which features of consciousness play this role and how they do it, I will leave aside for the time being. One thing I will say. It cannot be features that should be articulated in terms of first-person self-reference. If consciousness is part of what enables first-person self-reference, then the explanatorily relevant features of consciousness have better not be self-conscious thinking. Otherwise, the explanation ends up being explanatorily vacuous.

To sum up: Self-conscious thinking is a special form of thinking where the thinker refers to herself in a first-person way (with an “I”-concept). Four intuitive features characterize this form of thinking: three features concern the nature of first-person self-reference and one feature concerns its enabling relation to consciousness.

- (1) Semantic feature: guaranteed reference.
- (2) Epistemic feature: understanding and knowing that one is necessarily referring to oneself.
- (3) Motivational feature: first-person self-reference relates the thinker to the content of the thought in a special way.
- (4) Consciousness: phenomenal consciousness is part of what explains a person’s ability to self-refer in the first-person way.

The first three features and the last one are related to each other in a way that has a bearing on one’s theory of consciousness. If we accept it as a datum that self-conscious thinking is characterized by 1–3 and that part of what explains



our ability to refer to ourselves in this way are certain features of phenomenal consciousness, then we can derive important constraints for a theory of consciousness. From time to time, we think about ourselves in a way that is characterized by 1–3. This we cannot deny. If this ability is (partly) explained by phenomenal consciousness, then a viable theory of consciousness must be such as to allow for such an explanatory relation between consciousness and first-person self-reference in thinking. If a theory of consciousness cannot make sense of this explanatory relation, then our ability will be unintelligible.

In the following I will argue that any theory that denies that consciousness entails minimal self-consciousness, i.e. all anonymity theories, renders first-person self-reference unintelligible because they cannot make sense of any explanatory role for consciousness in first-person thinking.

### 3. Anonymity Theory of Consciousness

In what follows, I will call any theory denying the existence of minimal self-awareness an *anonymity theory of phenomenal consciousness* (anonymity theory for short). One way to grasp the anonymity theory is in the following way. According to the theory of minimal self-awareness, conscious experience does not only involve conscious representation or presentation of something (things, events, or facts), it also involves a conscious manifestation for a subject. Conscious experience is, as McGinn once remarked, Janus-faced: directed outwards as a presentation of the world and directed inwards as a presentation for me (1988). According to the minimal self-awareness theory, this for-me-ness aspect of conscious experience is part of what is consciously experienced. The for-me-ness is a phenomenal feature of the conscious experience. It is exactly this idea which is denied by the anonymity theory. The latter is not denying that experience, or more generally mental states and events occur to or belong to subjects. The fact that experience necessarily happen to a subject and in that sense belong to the subject is a conceptual and/or metaphysical fact. Rather the objection is that ownership of experience rather than being a feature internal to experience, is a purely external metaphysical relation (Tye 1995, 10–12, cf. Searle 2005).

In contemporary philosophy of mind, people defending so-called strong intentionalism seem to endorse a version of the anonymity theory. Strong intentionalism can be defined as the view which endorses the following two claims:

- (1) All states that are phenomenally conscious are intentional states (mental states with representational content); and
- (2) The phenomenal character of conscious states is identical to or supervenes on the state's intentional content (for discussion, see Crane 2009).

To see the intuitive appeal of this position, take the case of vision. When I open my eyes and enjoy visual experiences (i.e. look around), the world seems to present itself to me. I do not see representations of the world or fleeting phenomenal data – I see the objects themselves as they are located in front of me (say, on my desk). Capitalizing on this basic fact, we might say that all that is phenomenally present to me in a visual experience is the world as it is presented to me. The phenomenal properties that I experience are thus experienced as being properties of the objects that I am seeing. This becomes obvious, according to proponents of strong intentionalism, when I turn my attention introspectively to what it is like for me to have the experience. Every time I attend to features of my experience or its particular character I end up attending to features of the objects I am looking at. The experience is, in this sense, said to be transparent. (Dretske 1995, 2003a, Tye 1995, 2003).

Strong intentionalism is not merely a theory about visual experience. It purports to be a general theory of consciousness. The strategy proposed by the strong intentionalist is therefore to generalize the account of visual experience to all forms of conscious experience. Consequently, strong intentionalism is not the (weaker) view that all phenomenal states are intentional (Crane 2003). This kind of weak intentionalism could allow for phenomenal properties that are not identical to or supervene on properties of the content. For example, weak intentionalism would be consistent with the existence of phenomenal properties which are properties of the experience (for example, non-representational properties characterizing an experience as a visual experience, a recollection, or an imagining – in other words, characterizing the attitude and not the content). By contrast, the strong intentionalist claims that the phenomenal character of all forms of conscious state is identical to or supervenes on the state's intentional content. This means that conscious experience does not provide the subject with any direct experiential access to the experience itself. Hence, knowledge of being engaged in the conscious act of seeing (rather than imagining) has to be inferred from the way in which the world is being represented (Dretske 1999, 2003b, Byrne 2005, 2010).

One way to understand the idea of a completely anonymous or impersonal

experience is by analogy with cinematic representation. In his famous paper on imagination, Bernard Williams (1973, 37) pointed out that there are two ways in which we can understand perspective in cinematic representation. What is filmed is filmed from a location in space. The representation has lines of vision pointing back and converging in the perspectival location of the camera, and as a consequence things will be represented as appearing to the left, to the right, in the centre, in front of, being partly occluded, etc., with respect to this perspectival location or origin. Sometimes this perspectival origin is presented as belonging to a character in the represented world. If that is the case, then if the character scratches her nose, we will see her hand entering the picture, getting nearer the centre, and growing bigger in characteristic ways. Most often, however, the perspective is relative to no one. It does not belong to the camera, or to the director, or to me as a member of the audience. It belongs to neither of these since neither of these are part of the represented world. When we see two lovers kiss in a movie, often neither the camera, nor the director, nor the audience are parts of the world of the lovers. There is a perspective, but it is impersonal. There is nothing inherently different in the representational character or content with respect to the two forms of cinematic representation. In the first form of cinematic perspective, the perspective belongs to a character in the represented world by way of conventional signs. But in fact, both forms of cinematic perspective are equally impersonal.

If experience is strictly impersonal in this sense, then conscious experience does not in itself involve or imply any self-awareness when understood as this for-me-ness. For the same reason, conscious experience does not in itself ground self-knowledge: it provides the subject with immediate knowledge of the world, but not of her own conscious life. By contrast, according to the minimal self-awareness theory, experience not only presents the world but in a minimal, non-reflective way the experience also presents itself to the subject. Experience itself can be thought to directly provide the material needed for my self-conscious judgement "*I am looking* at a tree" in the sense that experience is providing an element which is necessary for knowing that I am engaged in the perceptual experience of focused looking. Not so for the anonymity theory. On the completely impersonal account of phenomenal consciousness, the phenomenal content of experience does not by itself provide the subject with any such element. The experience provides the subject with information about the world and only derivatively (and given some additional assumptions) with information about the experience (Dretske 2003b).

## 4. The Challenge: Explaining First-Person Self-Reference

The challenge I want to pose to the anonymity theory is that it falls short of supplying us with a convincing account of the way in which consciousness (is part of what) enables first-person thinking. If the anonymity theory of consciousness were true, then first-person thinking would be impossible or utterly unintelligible (mysterious). Since first-person thinking is neither, the anonymity theory must be false. But why think that the anonymity theory is inconsistent with the idea that consciousness plays an enabling role for first-person self-reference. There are at least two different ways in which consciousness as conceived by the anonymity theory could enable self-conscious thinking.

### 4.1. 1<sup>st</sup> Explanation: “I” as a Perceptual Demonstrative

Thoughts containing indexical constituents like *I*, *now*, or *here* are special. Their meaning can change from context to context. The thought *Now is the time to ask her* thought by me at time  $t_1$  has a different meaning than the thought *Now is the time to ask her* thought by me at time  $t_2$ . The thoughts have different truth-conditions. The meaning of these two thoughts changes with the context. The standard solution to such puzzles is to say that the meaning is bound to the context at least in the sense that a person can entertain an indexical thought only if suitably placed with respect to the objects being referred to by these thoughts (Heck 2002, 10). So, it is because the thinker is suitably placed vis-à-vis herself that she can think about herself in the first-person way.

One option for the anonymity theory would be to say that perceptual experience provides the subject of experience with an experientially based knowledge of herself. Perceptual experience thus places the first-person thinker in a suitable relation to the object of reference, viz. herself. This idea can be articulated in terms of the perspectival character of perception. The anonymity theory could use the perspectival nature of perceptual content as a form of uniquely self-specifying knowledge. In perception, things appear as located in space, to the right, to the left, in front, behind, within reach, out of reach, all with respect to my embodied point of view. Perceptual content thus has an implicit “back-reference” to the perceiving subject. If it did not, perception could not directly control action. The idea is not that the reference is fixed by what I am perceptually attending to (as with true demonstratives)

but rather by the indexical “back-reference” to the perceiver. James J. Gibson’s ecological optics and his notion of affordances might be exploited to substantiate this idea.

This idea could be used by the anonymity theory. One option would be to say that the implicit “self-reference” of perceptual content is what grounds our ability to think first-personal thoughts. The “I”-concept is a singular referring concept, and as such it might be thought that the use of it in self-conscious thinking requires that the thinker is in possession of knowledge that uniquely specifies the referent. The background assumption here could be that singular reference requires that the thinker knows which object she is referring to (Strawson 1959, Evans 1982, ch. 6, cf. Anscombe 1975, O’Brien 2007, ch. 2). The perspectival structure of perceptual content provides the perceiver with knowledge of a unique object: the self. This self-knowledge is what grounds the ability to refer to oneself in the first-person way. The indexical nature of the content of perception is what fixes the reference of the “I” and grounds our ability to think “I”-thoughts (Bermúdez 1998).

There are good reasons for thinking that this proposal does not work. First, there is the problem that uniquely self-specifying knowledge is perceptual whereas first-person thinking need not have any obvious relation to perception. It seems logically possible that a completely non-perceptual, purely “cerebral” creature could think self-conscious thoughts. If that is the case, then it would not be the structure of perceptual experience which fixes the reference of the “I” and provides the thinker with knowledge. A similar intuition was exploited by Anscombe in her famous thought experiment with the sensory deprivation tank (Anscombe 1975, O’Brien 2007: ch. 2).

Second, even if we disregard this first problem there will be reasons for doubting that the self-specifying knowledge provided by perceptual content is sufficient for making first-person self-reference possible. Remember that first-person self-reference in self-conscious thinking is such that a person cannot think an “I”-thought without knowing and understanding that she is referring to herself in the first-person way. On the account under consideration now, this knowing and understanding is explained by the perspectival nature of perceptual content. The problem for this idea is that we can imagine a person that loses this knowledge without losing her ability to first-person self-refer. Recall that the nature of the perceptual experience is supposedly impersonal in the sense that the cinematic representation is impersonal: it is referring back to some perspectival origin but the identity of the origin is left undetermined. The identity is picked out by the logical equivalent of a definite description: the perceiver of this perceptual presentation. As with any

definite description, we can imagine a situation where the thinker is referring to herself without knowing that this is what she is doing (Perry 1979). We can thus imagine a person enjoying or suffering some visual presentation (seeing a tree) who begins to wonder who the original perceiver actually is (i.e. who actually fills out the perspectival “origin-slot”). Such a subject might become convinced that it is in fact not her who is really seeing the tree but someone else (perhaps this chap Chris) who then flashes the perceptual presentations on to the subject’s mind as images on a screen. Paraphrasing a now famous statement by a schizophrenic patient reported by Frith (1992, 66), this subject might think: *I am seeing a tree but I am not really the one seeing it; this chap Chris is*. If something like this is possible, it goes to show that perceptual content cannot fix the reference of the “I” by perspectival “back-reference” and provide the thinker with the necessary form of self-knowledge.

#### 4.2. 2<sup>nd</sup> Explanation: Token-Reflexive Reference Rule

As an anonymity theorist, one might not be persuaded. One might think that the problem presented for the “demonstrative” account of “I”-thinking is a problem exactly for a “demonstrative” account and not for anonymity theory. To see this, recall the distinction made by Kaplan (1989a) between “pure indexicals” and “true demonstratives”. Pure indexicals are expressions like “I”, “my”, “today”, and “tomorrow”. According to Kaplan, the reference of pure indexical expressions uttered in a particular context is fixed automatically by simple reference rules. The reference of “I” would, for example, be fixed automatically by a rule like: a token of the “I” refers to whoever produces it.<sup>3</sup> By contrast, the reference of true demonstratives like “he”, “she”, “his”, “her”, “this”, and “that” is not fixed automatically in particular contexts. In the case of demonstratives, the reference depends on cognitive acts of demonstration by the utterer. As Kaplan writes, the reference in a context is fixed by demonstration and demonstration is “typically directed by the speaker’s intention to point at a perceived individual on whom he has focused” (1989b: p. 582).

We can transfer this insight from the domain of linguistic utterances to thoughts. Linguistic demonstration depends on mental demonstration which in turn depends on perceptual attention to particular objects. The “I”-con-

3 This idea is clearly expressed by Barwise and Perry (1981, 670): “Let us begin with the word ‘I’. A reasonable thing to say about this expression is that, whenever it is used by a speaker of English, it stands for, or designates that person. We think that this is all there is to know about the meaning of ‘I’ in English and that it serves as a paradigm rule for meaning”.

cept is not a demonstrative concept; it is the conceptual analogue of pure indexical expressions. The reference of the “I” is not fixed by an additional act of identification or demonstration. Reference is fixed automatically in the act of producing a token of the concept simply in virtue of the token being governed by the token-reflexive rule: a token of the “I”-concept is referring to its producer. This is all there is to first-person self-reference: the token-reflexive rule and the production of a token (of “I”) in a particular context. The anonymity theory might thus reply to the objections to the “demonstrative” account of first-person self-reference that the problems afflicting this type of account were to be expected given the fact that first-person self-reference (self-reference with an “I”-concept) is not demonstrative. First-person self-reference does not rest on an additional act of demonstration or identification (some form of self-knowledge); reference is automatically fixed by the token-reflexive rule governing the concept.

The anonymity theory can thus claim that all there is to the first-person self-reference of a self-conscious thought is that a constituent of the thought is governed by a token-reflexive rule. There are two versions of this view: a strong version and a weak version. The strong version has implausible consequences, whereas the weak version denies consciousness its explanatory role in first-person self-reference.

The strong version of the view is that first-person self-reference is exhaustively explained by the thinker’s mastery of the token-reflexive rule. “Mastery” means that the person is able to state the rule or at least is able to indicate when the rule is correctly applied and when it is not. The thinker understands that when thinking an “I”-thought she is necessarily thinking something about herself. The thinker has this understanding in virtue of consciously grasping the rule that a token of the “I” (in thought and talk) is referring to the producer of it, i.e., to herself. On this version of the view, consciousness plays a straight forward explanatory role. It is because of the thinker’s conscious grasp of the rule that she is able to engage in self-conscious thinking. Attacking this version is like attacking a straw-man. It is implausibly strong and to my knowledge has no advocates. The strong version over-intellectualizes first-person thinking and is most likely viciously circular.

The weak version of the view is to think of mastery of the rule purely in terms of behavioural dispositions. Here the anonymity theory understands the token-reflexive rule not as a rule that is explicitly represented by the thinker but as a law-like generalization that describes the workings of the system. The anonymity theory could claim that there is nothing more to first-person self-reference than the application of the token-reflexive rule in a

particular context of thinking (the rule-governed production of mental symbols). The rule need not be explicitly represented or grasped by the thinker. The rule is applied in the sense that it seems to govern the way in which the cognitive system functions. This view can provide a nice account of the cognitive difference between first-person and third-person self-reference. It could, for example, be claimed that the token-reflexive rule describes a situation where the cognitive system sets up a processing rule that either relates information directly to the cognitive system or does not do so. Application of the rule in a particular context describes a situation where certain information is put into an information file with direct pragmatic relevance to the system (Perry 1986). We would thus have functional story to tell about the motivational difference between the two forms of self-reference. This view also provides us with a story about guaranteed reference. It seems plausible that if a cognitive system produces a symbol to which token-reflexive rule finds application, then the symbol is guaranteed reference – after all, there must have been a producer of the symbol.

On this weak version, consciousness cannot play any explanatory role for first-person self-reference. It is a purely functionalist account of first-person thinking. All states and operations of the first-person self-referring system are defined by the input they take and the output they deliver. It seems very likely that a purely (analytical) functionalist theory is unable to explain or grasp phenomenal consciousness for the simple reason that we can always imagine that the same set of functionally defined states and operations could be realized in non-conscious machine or creature (Block 2007a).

Together with consciousness, one might think, goes the epistemic feature of first-person self-reference. Recall, the epistemic feature is the supposed fact that a person cannot think about herself in the first-person way without understanding and knowing that she is referring to herself. Again, it seems very likely that a purely functionalist theory is inconsistent with this particular intuitive feature. If we are interested in the special form of understanding that is manifested in self-conscious thinking, then the functionalist proposal seems insufficient. Thinking an “I”-thought entails thinking a thought where the thinker understands that she is necessarily thinking something about herself. It is this understanding which is missing from or left unexplained in the functionalist story. An equivalent of Searle’s Chinese Room argument might persuade us that this is so. Imagine a person inside the room sorting all incoming information into two files: information in one file is sent to the production system, whereas information in the other file simply stays there. Operating in a way that fits the rule does not entail that there is any



understanding. Or think of zombies. A zombie is functionally equivalent to a conscious human being and might, consequently, be operating in accordance with the token-reflexive rule. We might want to say the same of a computer or mindless robot. But we should refrain from saying that any of these creatures think self-conscious thoughts and understand that they are referring to themselves.

A natural response by the anonymity theory at this point would simply be to deny that consciousness, understanding, and knowledge play any significant role for first-person self-reference. First-person thinking is something we say occur when we have a cognitive system working in accordance with the token-reflexive rule. Self-conscious thinking is consequently something we have when we add phenomenal consciousness to a system that works in this way. By adding consciousness we might add a flavour but we add nothing of cognitive and explanatory significance.

I do not have any argument to show that this kind of theory is inconsistent; and I do not have any argument to demonstrate that the epistemic feature of first-person self-reference and the explanatory role of consciousness are undeniable. They strike me as intuitively necessary for first-person thinking. Denying them their role thus amounts to denying something which is intuitively correct. It therefore places an extra explanatory burden on the opponent, namely that of explaining why these ideas seem intuitively correct without actually being so.

The best I can do is to point to a number of contrast cases that make manifest the intuitive nature of the epistemic feature and the explanatory role of consciousness. The first case concerns the contrast between thinking and uttering. There is an important difference between thinking an “I”-thought and uttering an “I”-sentence. In the case of a person saying “I am not at home right now”, there is room for confusion and uncertainty about who is being referred to with the first-person pronoun. Just think of the message on my answering machine (spoken by a friend of mine) or the same message shouted out of my window by my slave when someone rings my door bell. When my slave shouts “I’m not in at the moment” we take him to be referring to me (Corazza, Fish & Gorvett 2002, Romdenh-Romluc 2006, 2008). Contrast such a case with my slave leaning out of the window thinking “I’m not in at the moment”. If we are attributing such a thought to him, it is clear that he could not be referring to anyone else but himself. We are attributing a state of understanding that he is making self-reference, otherwise we would attribute another type of thought to him (say, one where he is thinking about me). After all, when we are considering who the slave is referring to by his spoken

utterance, we are considering who he might be thinking of. In short, in the case of “I”-utterances there is room for referential ambiguity in a way there is not in “I”-thinking. One intuitive way to understand this difference is in terms of understanding. In the case of “I”-thoughts, we take the person to be understanding that she is referring to herself in the first-person way. In the case of the utterance, in considering whether the utterer is in fact referring to herself we are considering whether she understands that she is referring to herself in this way.

My claim is that this notion of understanding is explanatorily related to phenomenal consciousness. The second case concerns the contrast between conscious persons and non-conscious machines. Imagine that we have an intelligent computer that is able to pass the Turing test and that we know is not phenomenally conscious. It is to a very large extent functionally equivalent to us. When we ask it questions about what it can see and what it thinks about this or that, it will produce answers in the form “I can see...” or “I think that...”. In the Turing test conditions we would not be able to tell the machine apart from a self-conscious human being. In both cases we would automatically attribute self-conscious thinking to whoever is producing these answers. But on coming to know that the producer is in fact not conscious at all, it no longer feels right to attribute first-person thinking to it. An appropriate reply to this case by the anonymity theory is that, naturally, it does not feel right to attribute self-conscious thinking to a *non-conscious* machine; just like we would not attribute visual experience to a blind person.

Simply adding consciousness to the machine will not solve the problem. Consider a third case. Imagine a monoglot English speaker hypnotised to token the symbols “J’ai soif” in her subvocal inner speech. Let’s imagine that she is hypnotised in such a way that this string of symbols play a role which is functionally equivalent to a semantically equivalent string in her own language (“I’m thirsty”). Had she been given a glass of water, she would have drunk it, etc. She is producing an inner symbol governed by token-reflexive rule, she is conscious, and still she not thinking about herself in the first-person way.

I think the most intuitively appealing way to explain the difference in such cases between making conscious first-person self-reference and tokening of symbols that are governed by the reference rule is to say that what is missing in the case of the non-conscious machine and in the conscious hypnotised monoglot is understanding. In these cases, the machine and the hypnotised person is not producing the string of symbols understandingly. If a person thinks *I’m thirsty*, it means that she understands that she is referring to herself

in the first-person way. Consciousness is central to this notion of understanding. It is not just that, for some reason, only a conscious person understands; consciousness is in some way part of what explains our ability to understand. I don't know how to argue for these claims. They can certainly be denied. But with such a denial follows also a denial of an intuitively appealing way to explain the contrast between these cases.

To sum up the problems that first-person thinking is supposed to give the anonymity theory: There can be no first-person self-reference if the thinker does not understand and know that she is referring to herself in the first-person way. The intuitively most appealing way to understand the difference between understanding and not understanding attributes an explanatory role to consciousness. On the anonymity theory, consciousness cannot play this role and consequently be something that enables first-person self-reference. If these intuitions are correct and consciousness does play an explanatory role in enabling first-person self-reference, then the anonymity theory is false.

## 5. Minimal Self-Awareness and First-Person Self-Reference

Let us return to the initial argument of section 1. I have tried to sketch a number of reasons for accepting 4 and 5. To be sure, I have not provided anything like a real argument but, nevertheless, I have indicated that such an argument might exist. Let us assume for the sake of the subsequent story that 4 and 5 are true. If we accept the premises 1–3, it would follow that the theory of minimal self-awareness is true. If so, this theory should be expected to be able to explain how consciousness enables first-person self-reference.

We have learned a number of ways in which this enabling role cannot be articulated. Crucially, we should avoid thinking of consciousness as something that singles out and determines the reference of the “I”-concept, that is, as a uniquely self-specifying knowledge. First, it is hard to see how such a singling out of an object as oneself can take place without already involving a notion of the “I”. That is, without the person engaging in first-person self-reference. If that is the case, then consciousness as self-determining knowledge involves the very thing it is supposed to explain. Second, a mechanism for identifying objects of a particular kind can misfire. It could be malfunctioning in some way: pick out the wrong object or pick out no object. First-person self-reference is guaranteed success. We can therefore easily construct scenarios where the mechanism breaks down but the person can still think

about herself in the first person way (as long as we are not talking about a general mechanism for thinking or consciousness).

Then how should we account for the enabling role of consciousness in our first-person thinking? If first-person thinking entails understanding that you are thinking about yourself in the first-person way, then one attractive option is to tie the enabling role of consciousness to the understanding. A number of philosophers have recently argued that understanding is a particular type of cognitive experience. Here is Husserl's way of formulating what is meant:

Let us imagine that certain arabesques or figures have affected us aesthetically, and that we suddenly see that we are dealing with symbols or verbal signs. In what does this difference consist? or let us take the case of an attentive man hearing some totally strange word as a sound complex without even dreaming it is a word, and compare this with the case of the same man afterwards hearing the word, in the course of a conversation, and now acquainted with its meaning [...]. What in general is the surplus element distinguishing the understanding of a symbolically functioning expression from the uncomprehended verbal sound? (Husserl 2001: p. 105)

Let us follow Galen Strawson in calling the distinguishing surplus the “understanding-experience” (Strawson 1994: p. 5–13), a certain phenomenal character constitutively associated with the way in which the understanding person “automatically and involuntarily *takes* the sounds *as* signs, and indeed as words and sentences, that he automatically and involuntarily understands as *expressing certain propositions* and as representing reality as constituted in certain ways” (Strawson 1994: p. 6, italics in the original). I think what is important here is that this understanding-experience not only captures the way in which content changes between the two situations but also the way in which the attitude changes as well. There is something special about the “understanding-attitude”, the way in which content (as understood) is grasped.<sup>4</sup>

My proposal is that this cognitive phenomenology or understanding-experience can provide us with the experiential element that is missing in the purely functionalist account. A person genuinely thinking an “I”-thought is a person who is thinking the thought while phenomenally-consciously understanding it. Neither a machine, nor a hypnotized human who is internally (as a “thought”) or externally (as an “utterance”) tokening the sentence “*j'ai soif*” will do so while phenomenally-consciously understanding it.

4 For discussion of this kind of experience, see (besides Husserl and Strawson) Gallagher and Zahavi 2008, pp. 115–116, Horgan and Tienson 2002, and Pitt 2004. The most comprehensive and careful treatment of the “understanding-experience” is probably to be found in Ingarden 1968.

This line of reasoning gives us a sense in which first-person reference in thinking can be said to be grounded on phenomenal consciousness without claiming that phenomenal consciousness is a sort of reference-fixing self-knowledge. In order for a creature to self-refer in thinking she not only needs to be in possession of the right conceptual capacities, she also needs to be thinking the thought understandingly. To be thinking the thought understandingly is to have an understanding-experience. It is to be phenomenally-consciously grasping a proposition concerning oneself in a first-personal way.

There is another reason for thinking that the token-reflexive rule must be supplied with an additional phenomenal element in order to explain first-person self-reference. The token reflexive rule says that a token of "I" refers to whoever produces it. Imagine that the one of my thoughts is governed by this rule. Furthermore, assume that in order for this thought to be a genuine "I"-thought I must understand that I am referring to myself. I (my cognitive system) grasp the rule that one of the elements in this internal string of concept-tokens is governed by the rule. One of the elements refers to its producer. How do I (my cognitive system) know that I am the producer? In some cases we might be able to say that it was because I produced my thought intentionally and that I therefore knew because I intended to produce it (agent's knowledge). But I do not intend to produce all thoughts occurring in my mind. Some thoughts simply occur to me, and some of those thoughts are "I"-thoughts. One suggestion would be that I know I am the producer because when thinking "I"-thoughts, as when thinking all other thoughts, I have a certain conscious experience. I live through an episode characterised by cognitive phenomenology.

What is to stop the anonymity theory from adopting this idea of cognitive phenomenology? According to the anonymity theory, all phenomenal properties are properties of the objects as they are represented in the content. In so far as anonymity theory is consistent with the idea of an understanding-experience, it has to say that all the phenomenal features of the understanding-experience are features of the content, that is, features of the content being grasped or understood.

I want to highlight two problems for this suggestion. First, it seems to misdescribe the situation referred to by Husserl and Strawson. When the person in Husserl's example changes from consciously looking at a decorative arabesque to reading a word, this cannot be exhaustively described as a change in content of the experience. Certain aspects of the content of the experience do not change. The objects the person is looking at are the same in the two conditions. Certain other aspects of the content changes

dramatically between the two conditions: in the first condition the object is the decorative figure visually present in front of the person; in the second condition the object might be some proposition or state of affair and the figure visually present in front of the person can “shrink” into an almost transparent mediator. Connected to this change in content is a change in attitude or mode of consciousness. There is a change from *simply looking* at some decorative figure to a *taking* the figures as signs and *grasping* their meaning. This change of attitude or mode of consciousness is as much a part of the phenomenological story as the contents. To omit the attitude-part is to omit a significant part of the story.

Second, if the attitude or mode side could in fact without any residue be reduced to the content side, then the understanding-experience could no longer be said to enable first-personal self-reference in thinking. If the understanding-experience is to enable first-personal self-reference, then it must at least be something that necessarily is there whenever a person self-refers in first-person thoughts and without which the person cannot be said to understand herself as referring to herself. But it seems likely that a subject can always become alienated from the content of her experience in a way that is not intelligible if it is supposed to explain first-person self-reference. In this respect the anonymous version of the understanding-experience is no different from the visual experience discussed above. There will be nothing to tie the content necessarily to the subject. The following version of the experience would therefore be possible: some grasping of content is going on, but am I the one grasping the content? In other words, there might be some content which appear as “understood”, but I might doubt that this understood content was actually thought and understood by me. The point here is that even if ownership of understanding (parallel to perspectival ownership) is compromised, I would still be able to self-referringly use the “I”. Therefore, it could not be content alone which grounds and determines self-reference. Phenomenal content on its own would not be something which renders my use of the “I” intelligible to me.

Could the anonymity theory not respond that there is something absurd about the idea I could enjoy or suffer some phenomenal content which is not phenomenally mine? Could the anonymity theory not claim that if it is experienced by me, then it is *ipso facto* mine? There are at least two ways of understanding this response. First, it could be understood as meaning that any experience is the experience of a subject. In that sense, if there is an experience, then it belongs to a subject. Understood in this way, this belonging or ownership is an external metaphysical relation: experiences are such that,

necessarily, an experience is the experience of a subject. One way to spell this out would be to say that token experiences are events and, as Tye (1997, p. 332) says, “events are individuated in part via the objects which undergo them.” Experiences are on a par with explosions, killings, and screams. Just like a particular explosion is the explosion of some particular explosive thing, the experience is the experience of a particular subject. This belonging-to-a-subject or ownership is not phenomenal in any sense. But if that is the case, then even though this ownership might be true as metaphysical fact, it could not be something that necessarily enters into the understanding-experience. It would not be something we could allude to in order to explain the subject’s understanding.

Second, this mineness or ownership might be understood in a phenomenal sense. If this is how we understand it, then anonymity theory is committed to make it part of what is represented in the content of the experience. Mineness as part of the content, what could that be? A demonstrative or a description picking me out? Of course not. Whatever representation we can come up with it would seem that it could always be possible for the subject to doubt that she is the one represented in the content of the experience. So, it seems that the only option left is to accept that the mineness or ownership is a phenomenal feature outside the content. I think the only sensible suggestion would then be that it is a feature of the attitude or mode of experience. Accepting this suggestion is to accept a theory of minimal self-awareness. In other words, if first-personal self-reference in thinking is grounded on phenomenal consciousness (the “understanding-experience”) which makes self-reference understandable to the self-referring subject herself, then this phenomenality is not reducible to features of the representational content. The phenomenal residue is the phenomenal way in which the mode of consciousness is manifest to the conscious subject. These are phenomenal features of the experience or of the conscious subject. They are strictly first-personal. Experience cannot be completely anonymous. Therefore, anonymity theory is false and the theory of minimal self-awareness is true. Again, let me emphasize that these speculations falls short of providing us with a genuine argument. Still, I hope to have provided good reasons for thinking that such an argument exists.

## References

- Anscombe, G.E.M. 1975. The First Person. In S. Guttenplan (ed.). *Mind and Language: Wolfson College Lectures 1974*. Oxford: Clarendon Press: pp. 45–64.
- Barwise, J. and Perry, J. 1981. Situations and Attitudes. *Journal of Philosophy*, 78, 11: pp. 668–691.
- Bermúdez, J.L. 1998. *The Paradox of Self-Consciousness*. Cambridge, Mass.: MIT Press.
- Block, N. 2007a. Troubles with Functionalism. In N. Block, *Consciousness, Function, and Representation*. Cambridge, Mass.: MIT Press: pp. 63–102.
- Block, N. 2007b. On a Confusion about a Function of Consciousness. In N. Block, *Consciousness, Function, and Representation*. Cambridge, Mass.: MIT Press: pp. 159–214.
- Burge, T. 1998. Reason and the First-Person. In C. Wright, B. C. Smith, C. MacDonald (eds.). *Knowing One's Own Mind*. Oxford: Blackwell: pp. 243–270.
- Byrne, A. 2005. Introspection. *Philosophical Topics*, 33: pp. 79–104.
- . 2010. Recollection, perception, imagination. *Philosophical Studies*, 148: pp. 15–26.
- Cassam, Q. 1997. *Self and World*. Oxford: Clarendon Press.
- Castañeda, H.-N. 1967. On the Logic of Self-Knowledge. *Noûs*, 1: pp. 1, 9–21.
- . 1968. On the Phenomeno-Logic of the I. *Proceedings of the XIVth International Congress of Philosophy III*: pp. 260–266.
- Chalmers, D.J. 1996. *The Conscious Mind: In Search of a Fundamental Theory*. New York: Oxford University Press.
- Corazza, E., Fish, W. & Gorvett, J. 2002. Who is I? *Philosophical Studies*, 107: pp. 1–21.
- Crane, T. 2003. The Intentional Structure of Consciousness. In: Q. Smith & A. Jokic (eds.). *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- . 2009. Intentionalism. In: A. Beckermann, B. P. McLaughlin, and S. Walter (eds.). *The Oxford handbook of philosophy of mind*. Oxford: Clarendon Press: pp. 474–493.
- Dretske, F. 1995. *Naturalizing the mind*. Cambridge, Mass.: MIT Press.
- . 1999. The Mind's Awareness of Itself. *Philosophical Studies*, 95: pp. 1–2, 103–24.
- . 2003a. Experience as Representation. *Philosophical Issues*, 13 (1). pp. 67–82.
- . 2003b. How do you know you are not a zombie? In B. Gertler (ed.). *Privileged access*. Aldershot, England: Ashgate.
- Evans, G. 1982. *The Varieties of Reference* (edited by J. McDowell). Oxford: Oxford University Press.
- Flanagan, O. 1992. *Consciousness Reconsidered*. Cambridge, Mass.: MIT Press.
- Frankfurt, H. G. 1988. Identification and Wholeheartedness. In: H. G. Frankfurt, *The importance of what we care about*. Cambridge: Cambridge University Press, pp. 159–176.
- Frith, C. D. 1992. *The Cognitive Neuropsychology of Schizophrenia*. Hove: Lawrence Erlbaum.
- Gallagher, S. and Zahavi, D. (2008). *The Phenomenological Mind*. London: Routledge.
- Goldman, A. I. 1970. *A Theory of Human Action*. Princeton, NJ: Princeton University Press.
- Hech, Jr., R. G. 2002. Do Demonstratives Have Senses?, *Philosophers' Imprint*, 2, 2.



- Horgan, T. and Tienson, J. 2002. The Intentionality of Phenomenology and the Phenomenology of Intentionality. In D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings*. Oxford: Oxford University Press: pp. 520–532.
- Howell, R. J. 2007. Immunity to Error and Subjectivity. *Canadian Journal of Philosophy*, 37: pp. 4, 581–604.
- Husserl, E. 2001. *Logical Investigations I-III*. Trans. J. N. Findlay. London: Routledge.
- Ingarden, R. 1968. *Vom Erkennen des literarischen Kunstwerks*. Tübingen: Max Niemeyer.
- Kaplan, D. 1989a. Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives. In J. Almog, J. John and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press: pp. 481–563.
- . 1989b. Afterthoughts. In J. Almog, J. John and H. Wettstein (eds.), *Themes from Kaplan*. New York: Oxford University Press: pp. 565–614.
- Kriegel, U. 2009. *Subjective Consciousness: A Self-Representational Theory*. Oxford: Oxford University Press.
- McGinn, C. 1988. Consciousness and Content. *Proceedings of the British Academy*, 76: pp. 219–239.
- O'Brien, L. 2007. *Self-Knowing Agents*. Oxford: Oxford University Press.
- Perry, J. 1977. Frege on Demonstratives. *The Philosophical Review*, 86, 4: pp. 474–97.
- . 1979. The Problem of the Essential Indexical, *Nous*, 13: pp. 3–21.
- . (1986). Perception, Action, and the Structure of Believing. In R. E. Grandy and R. Warner (eds.), *Philosophical Grounds of Rationality. Intentions, Categories, Ends*. Oxford: Clarendon Press: pp. 333–361.
- Pitt, D. 2004. The Phenomenology of Cognition, or, What is It Like to Think That P? *Philosophy and Phenomenological Research*, 69 (1): pp. 1–36.
- Romdehn-Romluc, K. 2006. 'I'. *Philosophical Studies*, 128 (2): 257–83.
- . 2008. First-person thought and the use of 'I'. *Synthese*, 163 (2): 145–56.
- Searle, J. R. 2005. The Self as a problem in philosophy and neurobiology. In T. E. Feinberg & J. P. Keenan (eds.), *The Lost Self: Pathologies of Brain and Identity*. Oxford: Oxford University Press: pp. 7–19.
- Shoemaker, S. 1968. Self-Reference and Self-Awareness. *Journal of Philosophy*, 65, 556–579.
- Siewert, C. P. 1998. *The Significance of Consciousness*. Princeton: Princeton University Press.
- Strawson, G. 1994. *Mental Reality*. Cambridge, Mass.: MIT Press.
- Strawson, P. F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Methuen, 1959.
- Tye, M. 1995. *Ten Problems of Consciousness*. Cambridge, Mass.: MIT Press.
- . 1997. A Representational Theory of Pains and Their Phenomenal Character. In N. Block, O. Flanagan, and G. Güzeldere (eds.), *The Nature of Consciousness*. Cambridge, Mass.: MIT Press: pp. 329–340.
- . 2003. *Consciousness and Persons*. Cambridge, Mass.: MIT Press.
- Williams, B. 1973 1966. *Imagination and the Self*. In *his Problems of the Self*. Cambridge: Cambridge University Press, pp. 26–45.
- Zahavi, D. 2005. *Subjectivity and Selfhood: Investigating the First-Person Perspective*. Cambridge, Mass.: MIT Press.