Self-Involving Representationalism: In favor of a Weak Interpretation of Hume's Observation

One of the most famous Hume's quotes regards the relation between perception and oneself:

Hume's Observation "I never can catch myself at any time without a perception, and never can observe any thing but the perception." Hume (1739, p.252)

Most people tend to agree with Hume's observation, but disagree on the metaphysical conclusions to be derived from it: do selves exist? Are selves mere "bundles of perception"? Do they in endure? In this paper I am not interested in these metaphysical questions but rather in phenomenological considerations; and in particular whether there is such a thing as consciousness of oneself as a self, that is, as a subject of the experience. Hume's observation, *qua* observation about our experiences, suggests that among the phenomenal qualities that constitute our conscious experiences we cannot find one that correspond to an experience of the self or subject of experience (something like an I-qualia) beyong the perceived qualities in the world, emotions and sensations. In the 6th chapter of *The Conscious Brain* (Prinz (2012a), see also Prinz (2012b)), Jesse Prinz distinguishes a weak and a strong phenomenological interpretation of Hume's observation. According to the weaker reading, the reply to the question on whether there is such a thing as consciousness of oneself as the subject of the experience might be positive, i.e. there are "qualities" that correspond to oneself, but they "are nothing above and beyond the qualities of perception, sensation, and emotion." (ibid. 144). The stronger thesis, on the other hand, holds that there is no such thing as consciousness of oneself as subject, "there are no I-qualia, whether reducible or not". This stronger thesis is endorsed by Prinz. My aim in this paper is to defend, *pace* Prinz, a weaker version of Hume's observation.

This paper is organized in two sections. In the first one I argue in favor of a weak reading of Hume's observation. The force of Prinz's arguments against some of his opponents, and in favor of a stronger reading, rests on the failure of his opponents to provide a reductive understanding of what he calls *I-qualia*. For this purpose I will offer a such a reductive account in the second section.

1 In favor of a weak reading of Hume's Observation.

Many have found an important a priori support of this form of self-consciousness—as a subject rather than as an object as I might appear in others consciousness—in Decartes's *cogito.* Decartes claimed that the fact that he-himself thinks, the famous "I think", was the only thing that he could not doubt and implyed that there is an I, a subject, that can be directly accessed in consciousness. But as Prinz notes Decartes' argument remain controversial: what is undubitable is that there is thinking rather than that there is a thinker (Lichtenberg, 1765-1799)

Further a priori support might be found in Kant's transcendental argument in the *Critique of Pure Reason*, where he argues, roughly, that in being conscious of the world we are also conscious of ourselves as subjects of these conscious episodes given the unity of consciousness (see Brook, 2008 for example for details). In reply, Prinz notes that even if Kant were right, it is not clear whether this transcendental self is supposed to be phenomenologically present in experience—and this is precisely what the defender of the strong reading of Hume's thesis rejects—and he objects, moreover, that if it is the unity of consciousness what has to be explained, there are several proposals in the empirical literature that attempt to explain how components of experience get bound together without appealing to a self (Treisman, 2003; Tononi and Edelman, 2000; OReilly et al, 2003; Prinz, 2012a, ch. 7). I do not want to evaluate the merits of this reply here, because I think that, even if Prinz were righty, this form of self-consciousness has been independently motivated by appealling to the subjective character of experience.

One promising way of facing the task of providing a compreghensive account of the phenomenal character of experience is a divide and conquer one (Kriegel, 2009; Levine, 2001) that begins by making a conceptual distinction between two components of phenomenal character—the qualitative character and the subjective character—and the two associated problems. The qualitative character is what distinguishes different kinds of experiences; for example, the kind of experience I have while looking at my red apple from the one I have while, say, looking at a golf course. On the other hand, a theory of subjective character abstracts from the particular ways having different experiences feel and concentrates on the problem of what makes it the case that having a conscious experience feels at all. Hence, the qualitative character is what makes a state the kind of phenomenally conscious state it is, and the subjective character what makes it a phenomenally conscious state at all (Kriegel, 2009).

Our folk psychology dictates that conscious states are those states one is aware of *oneself* as being in (Rosenthal (2005)). In having an experience, say as of a red apple one is not merely aware of the apple but also in some sense aware of oneself as being in certain relation to the apple. What our experience conveys—the content of our experience—is not merely that there is a red apple but rather that I am perceptually related to the red apple, something like 'I am seeing a red apple'. This transitivity principle has been used to back up, for

example, Higher-Order theories of consciousness.[1]

A similar conclusion follows from phenomenological observation. There seems to be a common phenomenology of *for me-ness* common to all our experiences (see Kriegel (2009) and also Block (2007); Levine (2001)). This element is recognized along the phenomenological tradition as "pre-reflective self-consciousness". A detailed characterization of such an element is offered by Gallagher and Zahavi (2006):

> There is something it is like to taste chocolate, and this is different from what it is like to remember what it is like to taste chocolate, or to smell vanilla, to run, to stand still, to feel envious, nervous, depressed or happy, or to entertain an abstract belief. Yet, at the same time, as I live through these differences, there is something experiential that is, in some sense, the same, namely, their distinct first-personal character. All the experiences are characterized by a quality of mineness or for-me-ness, the fact that it is I who am having these experiences. All the experiences are given (at least tacitly) as my experiences, as experiences I am undergoing or living through. All of this suggests that first-person experience presents me with an immediate and non-observational access to myself, and that consequently (phenomenal) consciousness consequently entails a (minimal) form of self-consciousness.

Prinz complaints that he is unable to find in his experience this *for me-ness* that Kriegel and the phenomenologists talk about. This kind of phenomenolgical disputes are very difficult to settle, something that Prinz himself recognizes. Facing this controversy, he acknowledges:

> I don't mean to be suggesting that we should do without introspection entirely. Studies of consciousness often depend heavily on first-person reports. Rather the claim is that introspection on it's own may not suffice. Those who claim to find an I in experience would do well to find some non-introspective convergent evidence. Perhaps a substantive theory of what the phenomenal I consists in, backed up by non-introspective evidence for whatever the theory postulates, can help the Humeans see that there is an I in experience, after all. (p.157)

Providing such a sustantive theory will be the purpose of the next section; but it is worth considering first Kriegel's proposal in this direction and the reasons offered by Prinz to reject it.

We need to provide a characterization of the self-awarenness that our folk psychology and the phenomenological observation dictate is constitutive of our experiences. For this purpose I will just assume, as it is often done, that we can understand any form of awareness as some form or other of representation

[1] See for example Amstrong (1968); Carruthers (2000); Gennaro (2012); Lycan (1996); Rosenthal (1997, 2005)

and focus on the kind of representation required: conscious states are self-representational.

Kriegel holds, following Brentano, that conscious states are those that do not just represent the world but also themselves. In reply, Prinz presents two problems for Kriegel's proposal. The first one is that there seems to be no room for this kind of self-representation in a naturalistic framework because states do not cause themselves and causal relations seem to play a fundamental role in these theories. Kriegel (2009), however, unpacks the idea of self-representation in a way clearly compatible with naturalistic theories of mental content introducing the notion of indirect content and making use of the mereological distinction between complexes and sums. Roughly, the difference between mereological sums and complexes is that the way parts are interconnected is not essential for the former but it is for the latter—so a molecule of water is a complex and not a mere mereological sum. Kriegel concludes that a phenomenally conscious state, M, is a complex state that has two states, M* and M+, as proper parts, such that M* represents M+ directly and M indirectly in virtue of representing one of its proper parts—the analogy would be a case in which I indirectly represent a house by directly representing the fachade.

Second, Prinz objects that it is unclear how this form of self-representation is supposed to account for the alleged *for me-ness*: "It's not clear why a perceptual state that also represents itself would give rise to a phenomenal I." (p. 147). And I think that Prinz is right in his critics at this point.

We have seen that in the considered framework in which consciousness is unpacked in representational terms, conscious states are those that are self-representational. However, the expression 'M is self-representational' can mean either i) M represents itself or ii) M represents oneself. This two different approaches correspond to what Sebastian (forthcoming) calls mental-state involving (MSI) and self-involving (SI) theories.[2] Kriegel endorses a MSI view but the problem, as Prinz correctly remarks, is that MSI is insufficient for explaining the subjective character of experience and the phenomenology of *for me-ness*: what phenomenological observation suggests is that conscious states are bout myself and it is not clear how being about themselves would account for this fact. But can we alternatively make sense of self-involving representationalism and of Hume's observation at the same time?

2 Self-Involving Representationalism

Both folk psychology and phenomenological observation suggest that the subject of experience enters the content of experience: it is in the nature of the experience that it's correctness conditions concern the subject that is having

[2] It is important to remark, as Sebastian does, that the distinctionbetween MSI and SI theories is orthogonal to the one between 'non-egological' and 'egological' theories (Gurwitsch, 1941). In representational terms, the discussion between SI and MSI theories is a semantic one; i.e. a discussion about the content of experience. The distinction between egological and non-egological theories,on the other hand, is a metasemantic discussion: a discussion about what is required to account for such content.

the experience as such. Consider a visual experience as of a red apple. What my experience reveals is that both the apple and myself are constitutive of the content of the experience (the content is SI in opposition to merely MSI). But there is an important difference in the way both elements are represented: the apple is represented as an object whereas I am represented as the subject of the experience.

Prinz admits that the defenders of the strong reading do not have to deny that we are conscious of ourselves *as objects*. In order to clarify his eliminativist position rejecting the former remarks he writes:

> There is an obvious sense in which conscious experience of a self in incontrovertible. Many of my experiences are experiences of things that take place inside my body. A sensation on my skin or an emotion is in me. If the body is part of the self, then surely I can experience myself. So the eliminativist about the phenomenal I cannot deny that the self (or part of the self) can be an object of conscious experience. There is a phenomenal me. The issue concerns the subject of conscious experience. When I experience a sensation in my body, there is an experience of me, but is there also an I—a subject who is having that experience? Is there any experience that corresponds to the "I' when I say, "I am experiencing X"? The eliminativist will say no. One might put this by saying that the eliminativist denies any consciousness of a self as a self, that is, serving as the subject of an experience, thought, or action. More precisely, the eliminativist says there is no component of an experience that has a special claim to being the experience of self such that that component is playing something like a subject role for the experience. There is nothing phenomenal that corresponds to the "I" in states that we would express using that word.

This distinction between consciousness of oneself as a subject and an object matches the one made by Wittgenstein (1958) between the two uses of 'I' (or 'my'): a use as an object and a use as a subject as in 'I have a broken arm' and 'I am in pain' respectively. Wittgenstein maintains that "The cases of the first category involved the recognition of a particular person, and there is in these cases the possibility of an error, or as I should rather put it: the possibility of an error has been provided for. On the other hand, there is no question of recognizing a person when I say I have toothache. To ask 'are you sure it is you who have pains?' would be nonsensical." (ibid. 66-67). Shoemaker (1968) has argued that the use as a subject is *immune to error through misidentification relative to the first-person pronoun*, that is to say "it cannot happen that I am mistaken in saying "I feel pain" because, although I do know of someone that feels pain, I am mistaken in thinking that person to be myself." (ibid. 567): in the use as a subject no identification is required.

If one wants to accept that there is such a thing as representation of oneself as an object, as Prinz is willing to admit, then denying that representation of oneself as a subject seems not an option because, representation of oneself as

an object requires identifying oneself with an object and as Shoemaker (1968) has argued "not every self-ascription could be grounded on an identification of a presented object as oneself" (ibid. p.561). The reason is that identifying something S as oneself requires either finding something to be true of S that one independently knows to be true of oneself or finding that S stands to onself in certain relationship that only oneself can stand to one. But, as Shoemaker notes, this would in turn require that there is some property or relationship that we already had ascribed to ourselves which is not grounded on the identification in question. This would require another identification; so, if we want to avoid a vicious regress then not all self-ascription can be grounded on an identification. Shoemaker further remarks that the identification of an object as oneself is accompanied with the possibility of misidentification and hence, that representation of oneself as subject cannot depend on identification if it is to be immune to error through misidentification.

When I look at a red apple I have an experience as of red. My experience conveys, in a non-conceptual manner, that I myself (Castañeda (1966)) am confronted with a red object. Whereas I might be wrong about what is presented I cannot be wrong about the fact that is to myself to whom is presented. The reason is that in having an experience I do not need to identify myself with any kind of entity and my experience is priori to any such identification. I might fail to know that I am XX and thereby not knowing that XX is confronted with any object. A characterization of the content of experience requires the so called essential indexical (Perry (1979)); the correctness conditions of phenomenally conscious states concern the very same individual that is undergoing the experience *as such*, as the subject of experience: the content of experience is de se content (Castañeda (1966); Chisholm (1981); Lewis (1979)). So, in having a conscious experiences I do not merely attribute certain properties to the object causing the experience, I attribute to myself the property of being presented with an object with certain properties (Lewis (1979)). If this is the case, we can perfectly make sense of of Hume's thesis under a weak reading, because a "perception" of a red apple is at the same time about the apple and about myself.[3] However, when I introspect I do not find any object corresponding to the self, for I myself am not represented as an object in my experience and why there is no quality corresponding to the self on top of the qualities of perception, sensation, and emotion.

In order to provide a satisfactory reply to Prinz's challenge of providing a substantive theory, we need to get clear more clear about the notion of *de se* content and how a mental state comes to have such a content: what takes it to

[3] This idea is, I think, advanced by Shoemaker (1968) when he writes:

> But it is plainly not the occurrence of self-awareness in this sense (as a subject) that has been denied by those philosophers who have denied that one is an object to oneself; e.g., it is not what Hume denied when he said: '"I can never catch myself at any time without a perception, and never can observe anything but the perception." What those philosophers have wanted to deny, and rightly so, is that this self-awareness is to be explained in a certain way.(p.562)

self-attribute a property.

2.1 Understanding *De Se* Content.

I like the view about mental content according to which the role of mental states is to distinguish between different possibilities (Stalnaker (1999)). Content of mental states, are ways of dividing the space of possibilities, a division that is typically taken to be among ways the world might be; what is relevant to the content is that it exclude certain possibilities, certain ways the world might be. For example, my belief that XX is writing a paper is true or false—correct or incorrect—depending on the way the actual world is; in other words, it distinguishes two ways the world might be, one in which XX is writing a paper and one in which (s)he is not. Such partitions are made by attributing properties to objects; i.e. by representing objects as having properties—like that of writing a paper to XX in the example.

When I have a state with *de se* content, its correctness condition do not merely concern the way the world might be but also myself, the subject that is in the state. Propositions understood as something that determines partitions among possible worlds seems not to be well suited to capture its content. If we accept that my belief that XX is writing a paper and my belief that I myself am writing a paper have different correctness conditions despite the fact that I am XX—because, for example, I might be disposed to act ina a certain way if I belief one but not the other (Perry (1979), see also Lewis (1979), cf. Stalnaker 2008)—then the way the world might be seems to be insuffient for capturing these conditions, for both cases demand the very same thing from the world, namely that XX is writing a paper.

We need rather centered wolds. If a possible world is a way the world might be, a centered world can be thought as a way the world might be *for an individual.* Centered worlds propositions do not just individuate a way the world could be, but also a certain logical position within this world. Partitions among centered worlds are not made by attribution of properties to objects but rather, according to Lewis (1979), by self-attribution of properties, where a self-attribution of a property is not reducible to a mere attribution of a property to any particular object (in other words, it is not enought to attribute a property to oneself—to represent oneself as an object): self-attribution of a property requires "ascription of properties to oneself under the relation of identity" (Lewis, 1979, p. 543).

To make sense of this proposal a model of what it takes to self-attribute a property, at least in the case of our experiences is required. This is the purpose of the next subsection.

2.2 Naturalizing *De Se* Content.

My experience represents myself in a particular way that we have characterized as representation as a subject: the correctness conditions of the experience concern the very same individual that is undergoing the experience as such. In

having an experience I self-attribute (I represent myself as having) a certain property. Furthermore, this self-attribution should better not involve identification if Shoemaker is right—cf. Rosenthal (2011b).

A centered world semantics presented in the previous section provides a framework for understantanding this manner of representation; it provides a semantic of de se content. A naturalistic theory requires, on top of that, an explanation of the relation that holds between the vehicle of representation and its content, call this kind of theory a "metasematic" theory of mental content.

In the literature we can find several naturalistic metasemantic theories of "propositional content", explanations of what it takes to attribute a property to an object. These theories typically appeal to a causal relation between the object and the vehicle of representation. The underlying idea is that mental states represent what causes them in *normal condition* (to make room for misrepresentation— we do not want a mental state to represent anything that can cause it). Now, "normal conditions" is a normative notion, which is not acceptable in a naturalistic framework and theories of mental content attempt to unpack such normativity in naturalistic compatible terms. Let me present teleosemantic theories as illustration of one of these theories, for I will make use of their insight in my elaboration of a metasemantic theory of de se content. According to these theories, a representing system is one that has the teleological function of indicating that such-and-such is the case (tracking information about such-and-such), being such-and-such its content;[4] where the teleological funtion of the state plays the role of unpacking this normativity deciding which one among all the things the state might track information about is the one represented.

Hence, attributing, say, the property of writing a paper to XX is a matter, according to these theories, of being in a state M that has the teleological function of indicating that XX is writing a paper. But this is insufficient for understanding self-attribution because, as we have seen, XX can attribute to XX the property of writting a paper without self-attributing this property— I can believe that XX is writing a paper without believing that I myself am writing a paper.

The first question that should be faced in the search of an understanding of self-attribution is what kind of entities are individuals in the claim that the correctness conditions of the experience concern the very same individual that is undergoing the experience. In a naturalistic framework, organisms are probably the best candidates for this. Organisms are prior to experiences; this does not mean that representations of organisms are prior to the experience nor that I have to recognize myself as being a certain organism—I do not have to identify myself with certain organism. Having an experience as of a red apple cannot be

[4] This oversimplistic example amounts to the claim that representational states represent what causes them in normal circumstances, where the normative notion 'normal circumstances' is unpacked by appealling to the teleological function of the state. It is intended to capture the insight of teleological theories. For further and different elaboration of on the details see Dretske (1988); Millikan (1984, 1989); Mossio et al (2009); Neander (1991); Schroeder (2004).

a matter of representing one privileged organism and representing the apple. In this case we would have two representations as an object, and it is unclear how, and Shoemaker has argued not possible that, I can come to identify myself with such an entity. We need to explain how the organism represents itself as having a certain property, how such a self-ascription is possible without identification.

Organisms are continuously changing entities that remain nonetheless as functional unities, as unique systems, during the organisms' life. A widespread view in biology holds that living organisms are self-maintaining systems. The notion of self-maintaining system has a long history in philosophy dating back to Aristotle (Godfrey-Smith (1994); McLaughlin (2001)). In contemporary science it was popularized by cyberneticians but more recently, after Ilya Prigogine won the Nobel Prize in 1977 for his work on dissipative structures and their role in thermodynamics, many scientists started to migrate from the cybernetic approach to the thermodynamic view of self-maintaining systems.

In a self-maintaining system, the dynamics of the system tend to maintain the inherent order; its organizational pattern appears without a central authority or external element imposing it through planning. This globally coherent pattern appears from the local interaction of the elements that make up the system. The organization is, in way, parallel, for all the elements act at the same time, and distributed, for no element is a coordinator.[5]

If organisms are self-maintaining systems it seem appealing to look for the mechanisms that guarantee the stability within the organism's boundaries as the mechanisms that ground the distinction between what is part of the system—the organism—and what is not, the distinction between what is me and what is not, and might also be justified by the phenomenological sense of unity of all my experiences as being present for the same individual or self.

An interesting proposal in this direction is Damasio's notion of proto-self. In his book, *The Feeling of What Happens*, Damasio (2000) presented the proto-self as a constitutive element of our experiences. According to Damasio:

> The proto-self is a coherent collection of neural patterns which map [represent], moment by moment, the state of a physical structure of the organism in its many dimensions...[t]hese structures are intimately involved in the process of regulating the state of the organism. (Damasio, 2000, p. 154)
>
> It is an integrated collection of separate neural patterns that map, moment by moment, the most stable aspects of the organism's physical structure. (Damasio, 2010, p. 190)

[5] A simple example of these self-maintaining systems is the flame of a candle. In the flame of a candle, the microscopic reactions of combustion give rise to a macroscopic pattern, the flame, which makes a crucial contribution to maintaining the microscopic chemical reaction by vaporizing wax, keeping the temperature above the combustion threshold, etc. The flame itself favors the conditions that enable it to work. This is an example of the minimal expression of self-maintenance, called 'dissipative structures': Dissipative structures are systems in which a huge number of microscopic elements adopt a global, macroscopic ordered pattern (a 'structure') in the presence of a specific flow of energy and matter in far-from-thermodynamic equilibrium (FFE) conditions. Mossio, Saborido and Moreno (2009, p. 822)

The proto-self does not just map the internal milieu (the extra-cellular fluid environment) but also, for example, the musculoskeletal and visceral musculature. I will make use of this proto-self in my elaboration of the conditions under which a mental state comes to have de se content—in the particular case of experiences. I think that we can offer an account of such de se content by characterizing a conscious state as a complex of two states that I will call 'proto-self' and 'the proto-qualitative state':

On the one hand, the proto-self is a brain structure that has the function of regulating the homeostasis of the organism. It regulates the internal environment and tends to maintain a stable, constant condition required by the self-maintaining system; the stability required for life.

On the other hand, the proto-qualitative state is another brain structure that has the function of indicating that such-and-such state of affairs obtain in the world—for example the function of indicating that there is a red apple.

Different phenomenally conscious states are constituted by different proto-qualitative states. Proto-qualitative states are not phenomenally conscious: the proto-qualitative state doesn't have the required content. The proto-self is not a phenomenally conscious state either. It is the interaction between both of them that gives rise to a phenomenally conscious mental state that has the function of indicating that the very same organism that the proto-self regulates is being affected by the object the proto-qualitative state represents.

When looking at the red apple in front of me I undergo a phenomenally conscious experience. My visual system will generate a representation of the properties of the apple; this is a proto-qualitative state (PQ). But this is, still, an unconscious representation.

On the other hand, the organism has a subsystem, the proto-self, that monitors and controls the homeodynamics of the organism. The proto-self represents the status of certain internal state like the extra-cellular fluid environment, the musculoskeletal structure and the visceral musculature. This latter representation is altered by the processing of the apple—changes in the retina or in the muscles that control the position of the eyeball, but also changes in the smooth musculature of the viscera, at various places of the body, corresponding to emotional responses, some of them innate. At the level of content, this interaction will explain why the content of the complex state constituted by the proto-self and the proto-qualitative state is de se. The function of this complex is not just to indicate (to tracks information about) an object with certain surface reflectance nor to indicate that such-and-such bodily state obtains, but to indicates that the very same system that is doing the representing (given that the PS is part of the complex) is in certain perceptual state. When the organism is in this complex state it does not attribute a property to another object but it attributes a property—that of being in certain perceptual state—to itself, to the very same organism that the proto-self regulates, "under the relation of identity": the system attributes the property to the very same system that is doing the representing without any need of an identification. This complex state represents that the organism itself (Castañeda (1966)) is presented with an object that PQ represents.

This model, although very different at the explanatory level—in terms of de se content—shares with Damasio's model most of the neural structures that would underlie the neural correlate of our own experiences. Prinz 2012 pp.153-156 criticizes Damasio's model. Independently on whether those critizisms make justice to Damasio's model it is worth considering why they do not apply *mutatis mutandi* to the model I have proposed here. Damasio maintains, roughly, that conscious states correspond to second-order maps of how the processing of perceptual information is modifying the body—as represented by the proto-self. Prinz presents Damasio's proposal as follows:

> Damasio speculates that the nervous system must keep track of these changes [changes in the body corresponding to the impringment on our senses of objects in our environment]; it must monitor how the world in affecting the body at all times. To do this, there must be "second-order maps"—representations of the first-order body representations as they undergo changes.(p. 153)

and then goes and wonder whether experiences of body are necessary and sufficient for self-awareness. This seems result from a misunderstanding of Damasio's proposal; although Damasio claims that 'bodily feeling' are constitutive of our experience, he only refers by 'feelings' to the unconscious representations of bodily changes in the proto-self and he does not claim that body experiences are constitutive of say an experience as of a red apple. Prinz seems to be suggesting that the neural correlates of *bodily conscious experiences* such as proprioception are part of the neural correlate of an experience as of a red apple as the following quote against the necessary condition suggests:

> If a loud noise startles me, for instance, I will feel a strong sudden response in my bfollowingody. Attention is drawn inward as a feel myself reacting to the sound. But there are many other cases where the experience of the body is less pronounced without any loss in the impression that I am the subject of my experiences. Consider intellectual exercises such as reflecting on philosophy, answering crossword puzzle questions, or performing calculations in your head. Such activities might engage the body but, intuitively, they need not. To take one simple case, imagine counting backwards from 100. It seems very plausible that, while doing this, one periodically loses conscious experience of the body for brief moments.

But conscious experience of the body is clearly not required for having the kind of experience we have when we count for example. Damasio does not seem to make this claim and I definitely do not. Moreover, Prinz present empirical evidence showing that tasks involving working memory tend to suppress activity in brain areas associated with emotion and bodily experience (Yun et al., 2010). He maintains that when this happens the task seems "seem no less agentic, no less self-involving." Two things should be noted in reply. The first one is that the model I have presented do not claim that the more activity we find in the proto-self the more "self-involving" the experience should be, rather that activity there

is required for the content to be self-involving—to be *de se*—something that does not seem to admit different grades. But what is more important, when one looks into the details of the study by Yun et al. we can see that they show that activity in the dorso lateral prefrontal cortex—related to working memory—negatively correlates to activity in the amygdala and the ventromedial prefrontal cortex, both related to emotional processes. However none of these areas is postulated as the neural correlate of the experience (see Damasio (2000, 2010); Sebastian (forthcoming)); neither as part of the proto-self—corresponding to several brain stem nucli including the tegmentum; the hypothalamus; the insular cortex and S2—nor the structures that make the interaction between the proto-self and the proto-qualitative state possible—which might include areas like the superior colliculi; the anterior cingulate; the thalamus and the medial parietal cortex.[6] Prinz also attacks the claim that bodily feelings are sufficient for self-awareness, but, as it should be clear, my proposal does not make such a claim: the activity of the proto-self is not sufficient for it. For these reasons I think that the proposal I have presented in this paper has nothing to fear from Prinz's critics to Damasio's one.

3 Conclusion

In this paper, I have argued, *pace* Prinz, in favor of a weak reading of Hume's thesis.

I have argued that both folk-psychology and phenomenological observation suggest that awareness of oneself as a subject is required to provide a satisfactory account of the subjective character of experience. I do not take these arguments to be uncontestable and one might try to resist both. My purpose in this paper was rather to face Prinz's challenge of offering "a substantive theory of what phenomenal I consist in". I have argued that the correctness conditions of our experience concern the very same subject that is undergoing the experience as such and I have provided a model of what it takes to make the required self-attribution; a model that is supported by our current biological theories and the neurological evidence.

This model makes justice of Hume's observation: in introspection we only

[6] See Damasio (2000, 2010); Laureys and Tononi (2008) for empirical evidence suggesting that these areas are constitutive of the neural correlate of our conscious experiences. In his most recent work Damasio (Damasio, 2010) includes the cingulate cortex as part of the proto-self, I do not think that this are is required. Damasio postulates this area because bilateral anterior lesion of the cingulate causes a condition known as akinetic mutism, that is described by Damasio as "suspended animation, internally as well as externally"(Damasio, 2000, p. 176), however akinetic mutism is usually characterized as a variant of minimally conscious states as suggested b the following observation by Laureys:

> The interpretation is complicated by the fact that, in the rare instance in which such patients recover, there is usually amnesia for the akinetic episode, as in the original case of Cairns, though one patient who eventually recovered reported that she remembered the questions posed by the doctor but did not see a reason to respond (Laureys and Tononi (2008, p. 395))

find the qualities of perception, sensation and emotion corresponding to what is represented as an object, that does not mean that there is not such a thing as consciousness of oneself as a subject, rather that it is not to be found as something beyond the experience of the world, our body and our emotions.

Furthermore, this empirically supported and falsifiable model for self-representation dispels worries about the plausibility of self-involving approaches and the need to postulate any form of obscure and transcendental self, while might be suitable of providing the required unity of consciousness demanded by philosophers like Kant.

References

Amstrong D (1968) A Materialist Theory of the Mind. London: Routledge

Block N (2007) Consciousness, accessibility, and the mesh between psychology and neuroscience. Behavioral and Brain Sciences 30:481–548

Brook A (2008) Kant's view of the mind and consciousness of self. http://plato.stanford.edu/entries/kant-mind/, URL http://plato.stanford.edu/entries/kant-mind/

Carruthers P (2000) Phenomenal Consciousness: A Naturalistic Theory, 1St edition edn. Cambridge University Press

Castañeda HN (1966) 'he': A study in the logic of self-consciousness. Ratio 8:130–157

Chisholm RM (1981) The First Person: An Essay on Reference and Intentionality. Minneapolis: University of Minnesota Press

Damasio A (2000) The Feeling of What Happens: Body and Emotion in the Making of Consciousness, 1st edn. Mariner Books

Damasio A (2010) Self Comes to Mind: Constructing the Conscious Brain, 1st edn. Pantheon

Dretske F (1988) Explaining Behavior: Reasons in a World of Causes. MIT Press

Gallagher S, Zahavi D (2006) Phenomenological approaches to self-consciousness. http://plato.stanford.edu/entries/self-consciousness-phenomenological/, URL http://plato.stanford.edu/entries/self-consciousness-phenomenological/

Gennaro R (2012) The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts. MIT Press

Gurwitsch A (1941) A non-egological conception of consciousness. Philosophy and Phenomenological Research 1:325–338

Hume D (1739) A Treatise of Human Nature. Oxford University Press, USA

Kriegel U (2009) Subjective Consciousness: A Self-Representational Theory. Oxford University Press, USA

Laureys S, Tononi G (2008) The Neurology of Consciousness: Cognitive Neuroscience and Neuropathology, 1st edn. Academic Press

Levine J (2001) Purple Haze: The Puzzle of Consciousness. Oxford University Press

Lewis D (1979) Attitudes de dicto and de se. Philosophical Review 88(4):513–543

Lichtenberg GC (1765-1799) The waste books. New York Review of Books.

Lycan WG (1996) Consciousness and Experience. The MIT Press

Millikan RG (1984) Language, Thought and Other Biological Categories. MIT Press

Millikan RG (1989) Biosemantics. Journal of Philosophy 86:281–97

Mossio M, Saborido C, Moreno A (2009) An organizational account of biological functions. British Journal for the Philosophy of Science 60(4):813–841

Neander K (1991) Functions as selected effects: The conceptual analyst's defense. Philosophy of Science 58(2 ,):168–184

OReilly R, Busby R, Soto R (2003) Three forms of binding and their neural substrates: Alternatives to temporal synchrony. In: Cleeremans A (ed) The unity of consciousness: Binding, integration, and dissociation, Oxford University Press

Perry J (1979) The problem of the essential indexical. Noûs 13:3–21

Prinz J (2012a) The Conscious Brain. Oxford University Press

Prinz J (2012b) Waiting for the self. In: JeeLoo Liu JP (ed) Consciousness and the Self: New Essays, Cambridge: Cambridge University Press

Rosenthal DM (1997) A theory of consciousness. In: Block N, Flanagan OJ, Guzeldere G (eds) The Nature of Consciousness, Mit Press

Rosenthal DM (2005) Consciousness and mind. Oxford University Press

Schroeder T (2004) New norms for teleosemantics. In: Clapin H (ed) Representation in Mind, Elsevier

Sebastian MA (forthcoming) Experiential awareness: Do you prefer it to me? Philosophical Topics

Shoemaker SS (1968) Self-reference and self-awareness. The Journal of Philosophy 65:555–567

Stalnaker RC (1999) Context and Content: Essays on Intentionality in Speech and Thought. Oxford University Press, USA

Tononi G, Edelman G (2000) Schizophrenia and the mechanisms of conscious integration. Brain Research Reviews 31(2):391–400

Treisman A (2003) Consciousness and perceptual binding. In: Cleeremans A (ed) The Unity of Consciousness, Oxford University Press

Wittgenstein L (1958) The Blue and Brown Books. New York: Oxford