



THEORIA. Revista de Teoría, Historia y  
Fundamentos de la Ciencia

ISSN: 0495-4548

theoria@ehu.es

Universidad del País Vasco/Euskal  
Herriko Unibertsitatea  
España

Sebastián, Miguel Ángel

Consciousness and Theory of Mind: a Common Theory?

THEORIA. Revista de Teoría, Historia y Fundamentos de la Ciencia, vol. 31, núm. 1,  
2016, pp. 73-89

Universidad del País Vasco/Euskal Herriko Unibertsitatea  
Donostia, España

Available in: <http://www.redalyc.org/articulo.oa?id=339744130005>

- How to cite
- Complete issue
- More information about this article
- Journal's homepage in redalyc.org

redalyc.org

Scientific Information System

Network of Scientific Journals from Latin America, the Caribbean, Spain and Portugal

Non-profit academic project, developed under the open access initiative

# Consciousness and Theory of Mind: a Common Theory?\*

Miguel Ángel SEBASTIÁN

Received: 04/03/2015

Final Version: 25/08/2015

BIBLID 0495-4548(2016)31:1p.73-89

DOI: 10.1387/theoria.14091

**ABSTRACT:** Many have argued that the difference between phenomenally conscious states and other kind of states lies in the implicit self-awareness that conscious states have. Higher-Order-Representationalist (HOR) theories attempt to explain such self-awareness by means of higher-order representation. Consciousness depends on our capacity to represent our own mental states: our Theory of Mind. It is generally agreed that such an ability can be decomposed into another two: mindreading and metacognition.

I will argue that consciousness cannot depend on mindreading. The tenability of HOR theories depends, therefore, on the relation between mindreading and metacognition. I analyze several views on such a relation and argue that none of them seem to be a plausible option for HOR theories.

**Keywords:** Consciousness, Self-awareness, Higher-Order Theories, Theory of Mind, Mindreading, Metacognition.

**RESUMEN:** Muchos han argumentado que la diferencia entre estados fenoménicamente conscientes y otros tipos de estados reside en la auto-conciencia implícita que muchos estados conscientes poseen. Las Teorías de la Representación de Orden Superior (HOR) pretenden explicar esa auto-conciencia mediante representaciones de orden superior. La conciencia depende de nuestra capacidad de representar nuestros propios estados mentales: nuestra Teoría de la Mente. Se acepta en general que esta capacidad puede descomponerse en otras dos: la lectura de mentes y la meta-cognición.

Argumentaré que la conciencia no puede depender de la lectura de mentes. La sostenibilidad de las teorías HOR depende, por tanto, de la relación entre la lectura de mentes y la meta-cognición. Analizo algunas concepciones de esa relación y argumento que ninguna de ellas parece ser una opción plausible para las teorías HOR.

**Palabras clave:** conciencia, auto-conciencia, teorías de orden superior, teoría de la mente, lectura de mentes, meta-cognición.

---

\* I am very grateful to Ryan DeChant, Esa Díaz-León, Marta Jorba, Manolo Martínez, David Pereplyotchik, Mirja Pérez de Calleja, David Pineda, David Rosenthal, Josefa Toribio and two anonymous referees for discussion and comments. A previous version of this paper was presented at the XIV TIF, the Sopha 2012 and the Tecuemepe. I am grateful to the participants in these events, especially to Axel Barceló, Ángeles Eraña, Eduardo García-Ramírez, Ekain Garmendia and Carlos Muñoz.

Financial support for this work was provided by the Committee for the University and Research of the Department of Innovation, Universities and Business of the Catalunya government and the European Social Fund. Also by the DGI, Spanish Government, research projects FF12009-11347 and MI-NECO-FFI2014-51811-P, the PAPIIT projects IA400615 and IN401315, and the UNAM postdoctoral fellowship program.

## 1. *Introduction*

Many of our mental states are phenomenally conscious. It feels a certain way, or borrowing Nagel's expression, there is something it is like to be in these states. Examples of phenomenally conscious states are those one typically undergoes while looking at the ocean or at a red apple, drinking a glass of scotch or a tomato juice, smelling coffee or the perfume of a lover, listening to the radio or a symphonic concert, or feeling pain or hunger.

A theory of consciousness has to explain the distinctive properties that phenomenally conscious states have and other kind of states lack. Many have thought that this difference rests upon the implicit self-awareness that conscious states have. Higher-Order Representational (HOR) theories<sup>1</sup> attempt to provide this kind of explanation.<sup>2</sup> HOR theorists share the idea that a conscious state is a state that I am aware of being in—states we are completely unaware of do not count as conscious states—and unpack such an awareness as one form or other of representation. So, according to these theories, phenomenally conscious states are those that are the objects of some kind of higher-order process or representation. There is something higher-order, a meta-state, in the case of phenomenal conscious mental states, which is lacking in the case of other kind of states. This higher-order state represents (is about/is directed on to) the first-order state, which thereby becomes phenomenally conscious. Hence, according to HOR theories, consciousness depends on our capacity to represent our own mental states: consciousness depends on our Theory of Mind.

A Theory of Mind, henceforth ToM, is the ability of humans to identify their own mental states and attribute mental states different from their own to others. It is generally agreed that such an ability can, at least conceptually, be decomposed into another two: mindreading and metacognition.

Human beings are able to entertain representations of other people's mental states thanks to their mindreading ability. We attribute beliefs, perceptions, feelings or desires to other people and predict and explain their behavior accordingly. But we also, frequently, attribute mental states to ourselves. This kind of first-person access to our mental states is usually called 'metacognition'.

My purpose in this paper is to present a problem for HOR theories of consciousness by showing that the claim that phenomenal consciousness depends on a ToM is not plausible. HOR theorists seem to be committed to the claim that phenomenal consciousness depends on metacognition. Besides, some philosophers have maintained that it depends on our mindreading capacities. In section 2, I will argue against this latter claim. If my objection is sound, the tenability of a HOR theory would depend on the relation between mindreading and metacognition. I analyze several views on such a relation in section 3 and argue that none of them seem to be a plausible option for HOR theories.

---

<sup>1</sup> See, for instance, Amström (1968); Carruthers (2000); Gennaro (2012); Lycan (1996); Rosenthal (1997); Rosenthal (2005).

<sup>2</sup> It is controversial how to better characterize such self-awareness and whether HOR theories can provide a satisfactory account of it—see for example Kriegel (2003, 2009), Sebastián (2012), Zahavi (2005, 2006). In this paper I would like to remain neutral on this topic and present an independent problem for HOR theories.

## 2. *Phenomenal Consciousness and Mindreading*

HOR theories of consciousness try to explain what is for a state to be conscious by means of an awareness of that state (Rosenthal 2012). Such awareness is then unpacked as some form or other of higher-order representation of it. HOR theories commonly claim that a conscious mental state is the object of a higher-order representation of some kind.

The kind of representation that is required by the theory makes a basic difference among HOR theories. The main concern is whether higher-order states are belief-like or perception-like. The former are called Higher-Order Thought (HOT) theories (Gennaro 1996; Gennaro 2012; Rosenthal 1997; Rosenthal 2005) and the latter Higher-Order Perception (HOP) or ‘inner-sense’ theories (Amstrong 1968; Carruthers 2000; Lycan 1996). According to the former theories, when I undergo a phenomenally conscious experience as of red I am in a mental state with certain content. Call this content ‘red\*’. For this mental state to be phenomenally conscious, there has to be, additionally, a higher-order thought targeting it, whose content is something like ‘I see red\*’. On the other hand, HOP theories maintain that what is required is a (quasi-) perceptual state directed on to the first-order one. A second point of disagreement is whether a given state is conscious in virtue of its being actually the target of a higher-order representation (Gennaro 2012; Rosenthal 1997; Rosenthal 2005) or by the disposition to raise a higher-order representation (Carruthers 2000). Carruthers’s dispositionalist theory is the initial target of this paper.<sup>3</sup>

As we have seen, according to HOR theories, beings lacking metacognition lack thereby phenomenal consciousness, because the ability to represent our own mental state is a necessary condition for consciousness.<sup>4</sup> Carruthers (2000) further claims that the ability of mindreading is required.<sup>5</sup>

<sup>3</sup> It is worth mentioning a distinction between what Block (2011) calls ‘ambitious’ and ‘moderate’ HOR theories. The former “purse to capture what-it-is-likeness”, they attempt to provide a theory of phenomenal consciousness; the latter purpose to capture other kind of consciousness, “consciousness in one sense of the term, higher-order consciousness”. Block thinks that Lycan’s theory is probably a moderate one. However, most HOR theorists clearly endorse an ambitious one, as it is clear in the case of Rosenthal (2005), Carruthers (2003) or Gennaro (2012) and explicitly stated by Rosenthal (2011) or Weisberg (2011). In this paper I am interested in phenomenal consciousness and hence considering only “ambitious” theories.

<sup>4</sup> Rosenthal has recently defended (2012) that metacognition and the postulated higher-order representation has little in common beyond the fact that they both postulate higher-order psychological states. He convincingly argues that not all metacognition is a higher-order representation of the postulated kind. However, he provides no argument against the claim that higher-order representation requires the kind of metacognition I mentioned.

<sup>5</sup> Some philosophers consider this to be a reason for rejecting these theories. They are too demanding, for they require a mindreading faculty and most animals and arguably human babies lack it—though mindreading abilities have been recently demonstrated in corvids and canids (Bugnyar and Heinrich 2006; Hare and Tomasello 2005; Stulp et al. 2009; Udell et al. 2008). I do not consider this last point to be a defeating one. Maybe animals and babies lack phenomenally conscious states after all. Although intuitively they undergo conscious experiences, I can only be sure that I do have conscious mental states and I have no serious doubts that so does the reader. I do not think that a theory that maintains that animals and babies are non-conscious is immediately wrong. Surely, when comparing alternative theories, one that doesn’t have this consequence is to be preferred.

The insight of Carruthers' theory is that phenomenal consciousness is a matter of how things seem to us. This, in turn, depends on there being a distinction between how things are and how things seem to us. This distinction is a product of our mindreading capacity. As stated by Carruthers, some of the first-order perceptual states acquire, at the same time, a higher-order content by virtue of its availability to the mindreading faculty combined with the truth of some version of 'consumer semantics' (Millikan 1984; Millikan 1989; Papineau 1993; Peacocke 1995). Very roughly, the main idea of consumer semantics is that the content of a mental state depends on the powers of the system which "consumes" that state; for instance, what a state represents will depend on the kind of inferences the cognitive system is prepared to make in the presence of that state. Certain mental states are recognized as mental representations by the mindreading faculty and it is in virtue of their availability to this faculty, as a consumer system, that the perceptual states in question acquire a dual content. These states are phenomenally conscious states:

It is in virtue of the availability of first-order perceptual contents to a mind-reading system which understands the is-seems distinction and/or contains recognitional concepts of experience, that all of those first-order contents are, at the same time, higher-order ones... Each phenomenally conscious experience has its distinctive form of subjectivity by virtue of acquiring a higher-order analog content which precisely mirrors, and represents as subjective, its first-order content. (ibid. p. 243)<sup>6</sup>

Our evolutionary ancestors would have had first-order representational concepts for many features of the environment (RED, GREEN, etc.); then the development of a mindreading faculty would have allowed them to build up an *is-seems* distinction: a distinction between how things are and how things seem to us. The concepts produced by the mindreading faculty could make use of first-order representations and these higher-order recognitional concepts (SEEMS RED, SEEMS GREEN, etc.) could have been generated in response to the very same perceptual data that gave rise to the first-order concepts. This way, each experience would, at the same time, be a representation of some state of the world (for example, a representation as of *red*) and a representation of the fact that we are undergoing just such an experience (a representation of *seems red*), through the consumer system that is the mindreading faculty. This new content, *seems red*, is a by-product of a mindreading faculty, which builds up the *is/seems distinction*. In having an experience as of red, besides there being a first order representation of *redness*, there is also second-order representation of *seeming-redness*.

---

<sup>6</sup> Carruthers draws the distinction between beliefs and perception by appealing to that between analog and digital content:

[Perceptions] while perhaps being imbued with concepts (often or always), contain representations more fine-grained than any concept; these representations are analog. (ibid. p. 133)

Nothing from what I will discuss in this paper hinges on the details of the distinction between digital and analog content.

Carruthers introduces an interesting proposal defending HOR theories of consciousness:

The explanation of phenomenal consciousness which I am putting forward, then, claims that it is because the content of C is available to two sets of consumers—first order conceptual reasoning systems, as well as a higher-order mind-reading faculty—that those contents actually (categorically) have dual representational status. (ibid. p. 246)

A conscious mental state has a double content (*is/seems*) due to its availability to these two systems. The second content, provided by the mindreading ability, plays the role of explaining the differential features of phenomenal consciousness. This proposal, while compelling, faces, I think, a serious objection.

### 2.1. *Phenomenal Consciousness does not depend on Mindreading*

My purpose in this section is to show the implausibility of a theory of consciousness according to which having a phenomenally conscious experience depends on our mindreading abilities.<sup>7</sup> The reason is that phenomenal consciousness is a necessary condition for the attribution to others of mental states that feel some way or other, which is part of our mindreading ability. This objection may be extended to any HOR theory that makes mindreading prior to metacognition. In other words, either metacognition is prior to, or an independent mechanism from, mindreading, or higher-order theories face serious problems. In the next section, I will argue that these last options are not very plausible assuming the truth of HOR theories. Let me focus here on the relation between consciousness and mindreading.

When we mindread we attribute mental states to others. That is to say, we represent others as having mental states. Such mental state attribution legitimates certain evaluations: it has correctness conditions. Imagine the following situation with two subjects Sy and Sn. Sn is at home, lying on the couch drinking a beer, when Sy arrives. After saying hello and leaving the wallet on the table, Sy goes to the kitchen and open the fridge. A moment later, Sy comes back to the living room, looks at Sn, takes the wallet and leaves the flat. Sn thinks that Sy is mad at her because she took the last beer and Sy wanted to have one. Sn also thinks that Sy left the flat to go to buy beers. In this episode, Sn explains Sy's behaviour by attributing her certain beliefs and desires. Such mental state attribution can be evaluated for correctness. Sn's belief that Sy wants a beer is true or correct if Sy has the desire to have a beer and incorrect or false otherwise (imagine that Sy was just checking in the fridge whether there are the ingredients that she need for the pie she wants to prepare and she doesn't feel at all like having a beer because she wants to continue working).

Are conscious experiences required for such a mindreading faculty? My opponent would argue that they are not; for example, along these lines: creatures can see objects in the environment and the response of other organisms to those objects and their properties in different circumstances. Different properties cause different responses in different creatures. On that basis, organisms (through evolution) can come to theorize that there are internal states inside of other creatures that track particular properties and conditions. This mindreading faculty allows Sn, for example, to attribute Sy the desire to have a beer.

<sup>7</sup> Carruthers (2009) seems to take feelings as inputs for a mindreading ability.

Consider now the case of an attribution of a perceptual state, like the one Sn attributes when Sy opens the fridge. Sn believes that Sy is seeing the empty fridge. According to my opponent, when people attribute to others sensory states, there is no reason for attributing feelings, we just attribute to them states that track certain properties. Sn would be attributing Sy a state that tracks certain properties of the fridge without any need to attribute her any phenomenally conscious experience.

This kind of mental state attribution is very different from the kind we often do, for we often attribute conscious experiences. Beyond such a common sense observation, empirical evidence for the claim that we do attribute phenomenally conscious states to others can be found, for example, in a study by Gray et al. (2007). The authors of this study distinguish two kinds of mental state attributions related to *Experience*—like hunger, fear, pain, pleasure, etc.—and others related to what they call ‘Agency’—which includes self-control, morality, memory and emotion recognition. Participants in the experiment compare several characters (e.g., God, a fetus, a frog, a complex robot, a corpse) on a range of mental capacities. For example, participants were asked “whether a girl of 5 is more or less likely to feel pain than is a chimpanzee” (ibid. p. 619). Interestingly for our purposes, they found a strong correlation between attributions of experiences together with a poor correlation with the other group. For instance, while a human baby scored low on Agency abilities, she scored high on experience abilities and a complex robot scored very low on having experiences, but had a moderate Agency. The correctness conditions of conscious experiences attribution seems to concern the phenomenal character of our own experiences. The question is, how can one ascribe others with mental states that feel a certain way for the subject if one has never been in a mental state that feels?<sup>8</sup> It seems to me that the kind of mental state attribution that a normal subject who hasn’t undergone the relevant experience does would be completely different in this case.<sup>9</sup> Let me illustrate the idea with an example.

Cases like pain or orgasm attribution are particularly illustrative for my case. Imagine that Sn has never had an orgasm in her life whereas Sy had. However, Sn and Sy seem to be equally good in recognizing, given their behavioral response, whether their partners are feeling pleasure in sex or just faking. One day, Sn and Sy decide to share the night with Sx. When Sx starts to moan, both Sn and Sy attribute her a mental state that they would

---

<sup>8</sup> Recall that I am using the expression ‘states that feel a certain way’ as synonymous of ‘states such that there is something it is like for the subject to be in’. Some defender of Higher-Order theories might object that there is a legitimate use of ‘feel’ according to which there are sensory and perhaps emotional states, which folk psychologically we describe as feelings (corresponding to what Rosenthal calls ‘qualitative states’) but such that there is nothing it is like to be in them and that I am, therefore, begging the question against her position. This would be a misunderstanding of the point I am trying to make here, because there is just a terminological disagreement about the use of the term ‘feeling’. If there is such a legitimate use—or is the correct one, which I doubt—, then the subject can have a feeling without this feeling being conscious in the phenomenal sense; i.e., in the sense of there being something it is like to be in this state (other forms of consciousness are not relevant in this dispute for recall that I am just considering “ambitious” HOR theories and therefore there is nothing it is like to be in a qualitative state; see fn. 3). My claim can be straightforwardly restated in terms of phenomenally conscious states, which is what it was meant by *states that feel a certain way*.

<sup>9</sup> This is independent of whether my ascription of mental states to myself or others is due to a simulation theory or purely theoretical. I will say more on the distinction between simulation theory and theory-theory in the next section.



express by saying “Sx is having an orgasm”. Are they attributing the same kind of mental state? They do not. To see why this is so, we have to look for the correctness conditions of their respective attributions and see whether they are the same ones. It seems obvious that in their attribution they both attribute a phenomenally conscious state: if being in Sx’s state doesn’t feel a certain (pleasurable) way, then their state attribution would be wrong. But, Sy is not merely attributing to Sx a state that feels (pleasurable) but a state that feels relevantly similar to her own orgasms: Sy attribution is correct if Sx’s state feels relevantly similar to hers and false otherwise.

This kind of mental state attribution is not possible in the case of Sn because she has never had an orgasm. To remark that this is the case, imagine that two weeks ago, Sn tasted a new gourmet chocolate; as a result she got goose pimples and she let a moan go. This was the most pleasurable experience she had never had and she comes to believe that orgasms must be similar experiences. In these circumstances, Sn might be attributing to Sx a mental state which is relevantly similar to her tasting chocolate experience. One day, Sn has an orgasm for the first time. Now, the kind of experience she will be attributing to others when having an orgasm is different from the one attributed before she felt an orgasm for that first time. She knows how it feels to have an orgasm and will attribute to others a similar sensation when they are having one. In particular, she will maintain that Sx was not undergoing the kind of episode she was attributing to her in the previous example.

This example suggests that the kind of mental state attributions that someone that lacks phenomenal consciousness can do, in case she can, are different from the ones that I can do. If this is right, then phenomenal consciousness cannot depend on mindreading capacities, for phenomenal consciousness is prior, at least to certain, mindreading capacities; in particular, to our capacity to attribute conscious experiences. We attribute to others phenomenally conscious mental states and this kind of attribution is not possible unless one has undergone the relevant experience, as the example suggests.

Before moving on to discuss the relation between mindreading a metacognition, let me make some clarifications to be clear about the scope and the assumptions of the argument and consider some possible objections.

## 2.2. *Some Clarifications and Objections*

There is something in my argument that might seem to resemble the reasoning in the well known Knowledge Argument (Jackson 1982). Mandik (2010) argues that the acceptance (explicit or not) of what he calls the ‘Experience Requirement’—the “thesis that, for some experiences at least, and red [orgasm] experiences in particular, knowledge of what it’s like to have such an experience requires that the knower has had or is currently having such an experience” (pp. 233-234; see also Papineau 1999; Stoljar 2005; Tye 1999)—is the most plausible support of Jackson’s knowledge argument in favor of an epistemic gap. If my argument depends upon the truth of this principle, it might be subject to the same objections that have been offered to the knowledge argument.

In reply it has to be noted, as Mandik does, that most materialist (what he dubs “gappy-materialist” or type-b materialist in Chalmers’ (2003b) terminology) accept an epistemic gap but deny the ontological one. Interestingly for current purposes, this form of materialism is held by most higher-order theorists like Carruthers’ (Carruthers and Veillet 2007); Rosenthal (2005), for example, denies that the transitivity principle—the principle that a



conscious state is one I am conscious of myself as being in—that backs up HOR theories is a conceptual necessity, and he thinks that it rather derives from our folk psychological notion of consciousness.<sup>10</sup> Consequently, it seems reasonable to assume that my opponent is happy to accept something along the lines of the Experience Requirement, for she does not endorse a form of a priori physicalism and this assumption is not problematic. However, my argument does not require the Experience Requirement to be true but rather something much weaker, for it is a claim about the way in which we make such attributions.<sup>11</sup>

The argument does not require a commitment to the claim that it is not possible to attribute the very same kind of mental state without having undergone the relevant experience. If, for example, experiences E were identical to such-and-such brain activity, by attributing you such-and-such brain activity—say by means of a neural image—I might be attributing you experience E, independently of whether I have undergone experience E. But such an attribution would not be done *by means of the machinery that constitutes my mindreading faculty* and therefore it is not in tension with the claim that attributions of experience by means of mindreading faculty require phenomenal consciousness. Moreover, I am also happy to concede the (conceptual or even nomological) possibility of what Mandik calls ‘deviant subjects’ (see also Dennet 2007), subject who know what it is like to see red despite not having undergone the relevant experience. Mandik argues against the Experience Requirement by considering Swamp Mary, a microphysical duplicate of post-experiential Mary, who, he claims, has phenomenal knowledge. I am happy to accept that deviant subjects can make mental state attributions in spite of the fact that they have not undergone any relevant experience. Again, my argument does not intend to show that having the experience is the only way we can gain phenomenal knowledge (and this is what the knowledge argument seems to require), just that this is the way we, ordinary subjects, acquire the knowledge required for our mindreading capacities. And the denial of this is a commitment of my opponent’s proposal, as I have shown.<sup>12</sup> In a nutshell, phenomenal consciousness cannot be a by-product of *our* mindreading capacities, precisely because *our* mindreading capacities require phenomenally conscious mental states.

Finally, let me stress that I am not denying that there are certain mental state attributions that an ordinary person can make in spite of the fact that she has never undergone the experience. Consider a born-blind person. If I utter ‘there is a red object in the table’, she will understand—at least to some extent—my utterance and probably attribute me some mental state. But color-sighted people can go beyond such attribution. Chalmers (2003) distinguishes different kinds of phenomenal concepts—concepts that refer to phenomenally conscious experiences. Consider the case of an red experience. He distinguishes *re-*

<sup>10</sup> For an argument against the claim that HOR theories are a priori true—i.e. that the transitivity principle holds with conceptual necessity see Byrne (2004).

<sup>11</sup> I am therefore not preventing an a priori reduction (not even an a posteriori reduction)—for it might perfectly be the case that consciousness can be conceptually reduced to other function or state or anything. It might even be the case that something along the lines of the transitivity principle is true, in so far as it can be unpacked in terms that do not require cognition of our own conscious experiences—see for example Kriegel 2009 and Sebastian 2012.

<sup>12</sup> One last point is worth stressing. Contrary to Jackson’s argument I am not speculating about the knowledge a science fiction scientist might have but rather considering ordinary subjects and the correctness conditions of the mental state attribution they make before and after having the experience.

*lational, demonstrative and pure phenomenal concepts.* Among relational concepts we can distinguish *community* and *individual relational concepts*. The former, in the case of an experience as of red, can roughly be glossed as “the phenomenal quality typically caused in normal subjects within my community by paradigmatic red things” and the latter as “the phenomenal quality typically caused in me by paradigmatic red things” (p. 228). Moreover, one can refer to her experience indexically deploying a demonstrative content one might express by saying “this kind of experience”. Finally, there are pure phenomenal concepts introduced by Chalmers as follows:

[C]onsider the knowledge that Mary gains when she learns for the first time what it is like to see red. She learns that seeing red has such-and-such quality in her, and in other members of her community. She learns (or gain the significant belief) that the experience she is now having has such-and-such quality, and that the quality she is now ostending is such-and-such. (p. 229)

The case of Mary is only illustrative here. As it should be clear by now, I do not want to commit myself to the claim that undergoing the experience is the only way one can acquire such concept (Swamp Mary might possess a pure phenomenal concept),<sup>13</sup> but merely to the weaker claim that we acquire pure phenomenal concepts after having undergone the relevant experience: a colour-blind person, at least typically, does not possess this concept.

It seems clear that a color-blind person does not make color experience attributions by means of an individual relational concept because there is no phenomenal quality typically caused in her by paradigmatic red things. Demonstrative concepts would not do the trick for my opponent either. The reason is that there is nothing that is demonstrated that grounds the deployment of the concept if there is no conscious experience. Now, the born-blind person can possess at the very least a community related concept and attribute mental states under such concept. The community relational concept is arguably, as Chalmers notes, what is deployed in the public-language. If the color-blind person can have this concept, one might think that one can attribute experiences under a community relational concept, and therefore that consciousness is not prior to our mindreading capacities. However, it is doubtful, to say the least, that community relational concepts are meaningful, and that they might be possessed if no subject in the community undergoes experiences, as my opponent is committed to maintain. More importantly, our mental mindreading abilities go beyond this. When I attribute my colleague a red experience I do not thereby merely deploy a community relational concept: if I were to discover that the person I am attributing the experience is color inverted with respect to me I would say that my attribution was wrong irrespectively on the kind of experience the rest of the community have in this circumstances.<sup>14</sup>

Once the assumptions of the argument are clearly spelled out, there seems to be a way in which one can resist the argument. One might agree that attributions of conscious experiences to others require having that kind of states, and, at the same time, maintain that phe-

<sup>13</sup> Mandik and Chalmers disagree on this point. I would like to remain neutral and endorse the weak reading I am presenting, for my argument does not depend on the details of their discussion.

<sup>14</sup> If one is not moved by inverted spectrum and is willing to resist its conceptual possibility, there is empirical evidence showing that normal subjects differ in the experience they undergo when looking at the very same object in the same situation (for a review see Block 2007. See also Nida-Rumellin 1996). This all that my reasoning requires.

nominal consciousness depends on mindreading, by distinguishing “two modes” of the mindreading faculty: one that relates to attribution of knowledge, desires, beliefs and perception and other that relates to attributions of experiences. The example above shows that some attributions of mental states depend upon having conscious states; so, one possible alternative theory, not clearly in the spirit of Carruthers’ one, would maintain that our mindreading faculty evolved in two steps. In a first step, one of the modes of the mindreading faculty would have evolved allowing our ancestors to attribute states with certain functional role (if we focus on perceptual states, states that track certain properties).<sup>15</sup> If the mindreading faculty derives from the ability to theorize about others behavioral propensities—more on this theory-theory view of mindreading below—, then our ancestors might have developed such abilities without any need of getting conscious experiences involved. We can call this kind of faculty a “proto-mindreading” faculty, because it does not yet allow the kind of mental attributions that we do. This proto-mindreading faculty can also be directed to our own behavioral propensities and so we come to represent ourselves as being in certain states; thereby being *aware* of such state, and the state becoming conscious: at this point there is something it is like to be in such a state.<sup>16</sup> A mental state becomes phenomenally conscious in virtue of being available to this proto-mindreading faculty. However, our mind reading abilities go beyond those of such proto-mindreading faculty, for we can also attribute phenomenally conscious mental states to others. So, in a second step a *full-blown* mindreading faculty would have evolved allowing the attributions of phenomenally conscious states to others.

The problem of this reply is that, according to Carruthers, the functional role that such a proto-mind reading faculty would attribute exhausts the phenomenal character of experience. The mental attributions that this proto-mindreading would allow, would be sufficient for explaining ours and other creatures’ behaviour. So, there is no evolutionary advantage in attributing phenomenally conscious mental states. However, new brain structures would be required to make such phenomenally conscious states available for our mindreading ability so that we can attribute conscious experiences to others, because the inputs for a full-blown mindreading faculty (one that allows us to attribute conscious experiences to others as ours does) are not merely behavioral propensities but also our own conscious experiences. The problem in turn is that brain tissue is metabolically expensive (Aiello and Wheeler 1995). Therefore, it seems reasonable to assume that the required structures would not be constructed and maintained unless they contribute in someway to the fitness of the individual. But there is no justification for the evolution of the required mechanisms underlying this new *full-blown* mindreading faculty: the proto-mindreading suffices for explaining ours and others behavior.

If this is right, then phenomenal consciousness is prior to our mindreading ability and not a by-product of it. The plausibility of a HOR theory will, therefore, depend on the relation of mindreading and metacognition. Such a relation is controversial.<sup>17</sup> In the next section, I review different models and argue that the plausibility of a model in which mindreading is not prior to metacognition is hardly compatible with the truth of HOR theories.

<sup>15</sup> Or states that satisfy some isomorphic relation with the properties in our environment, what Rosenthal calls “qualitative states”—see fn. 8.

<sup>16</sup> Recall that I am restricting my critics to “ambitious” theories of consciousness. See fn. 3.

<sup>17</sup> See Carruthers (2009); Carruthers (2011) for an excellent review.

### 3. *The Relation between Metacognition and Mindreading*

If one is interested in the relation between metacognition and mindreading there are two possibilities that one has to consider: either mindreading and metacognition involve independent mechanisms or they have a common architecture.

A model in which metacognition and mindreading are independent capacities realized by distinct cognitive mechanisms has been proposed by Nichols and Stich (2003). This model, however, is not compatible with HOR views on the nature of phenomenal consciousness. Defenders of HOR theories can deny that phenomenal consciousness depends on mindreading, conceding the intuition presented in the previous section, while urging that phenomenal consciousness depends on metacognition. This view is, however, incompatible with the view that mindreading and metacognition are independent to each other, as Nichols and Stich's model suggests. The reasons are that i) the previous example suggests that phenomenal consciousness is a necessary condition for our mindreading ability and that ii) according to HOR theories, metacognition is a necessary condition for phenomenal consciousness. These two premises entail the conclusion that metacognition is necessary for our mindreading ability and both abilities cannot, therefore, involve independent mechanisms.

Contrary to the proposal by Nichols and Stich, it is commonly held that there is a unique mechanism for both abilities and that they are directly connected. There is, however, a huge controversy on whether metacognition is prior to mindreading (where metacognition being prior to mindreading means that the ability of mindreading depends on the mechanisms that evolved for metacognition) or the other way around.<sup>18</sup>

Goldman (2006) suggests that metacognition is prior to mindreading. The attribution of mental states to others depends upon our introspective access to our own mental states together with processes of inference and simulation of various sorts, where a simulation is "the process of re-enacting or attempt to re-enact, other mental episodes." This is what is known as simulation theory of mind. An example by Goldman and Shanton (2010) may help to illustrate the idea:

Seated in my living room on a wintry day, I might imagine myself instead watching the surf on some sandy beach. What I am trying to do is undergo a visual experience that matches (as closely as possible) a visual experience I would have if I really were on the beach. Vision science tells us that what transpires in visual cortex when undergoing visual imagery can, to a considerable extent, match what goes on during genuine vision (Kosslyn and Thompson 2000). This is what we call a mental simulation. This is a case of intra-personal simulation: trying to re-enact an event in one's own mind. In using simulation to read others' minds, however, one would try to re-enact their mental states. That's just how mindreading characteristically takes place, according to simulation theory (ST).

---

<sup>18</sup> Strawson (1959) defended on a priori reasons that, at the very least, mindreading depends on metacognition. He wrote: "A necessary condition of one's ascribing states of consciousness, experiences to oneself, in the way one does, that one should also ascribe them, or be prepared to ascribe them, to others that are not oneself... One can ascribe states of consciousness to oneself only if one can ascribe them to others" (pp. 99-100).

Defender of the priority of metacognition to mindreading, like Goldman, as well as other participant in the debate, seem not to be moved by Strawson's arguments, which arguably rely on his views on persons; a view that they might not endorse. So, I will not take any stance on such a priori reasons. The kind of reply that I am considering here in favor of my opponent requires that a conceptual distinction between metacognition and mindreading is possible and that metacognition be prior to mindreading, *pace* Strawson.

The thesis that simulation plays a role in mindreading, especially “when reasoning about the conclusions of other people’s reasoning and decision-making processes, is now widely accepted. Indeed, almost all theorists, of whatever stripe (theorizing theorists and modularists included), now find a place for simulation within their accounts.” (Carruthers 2011, 225). The controversial claim of proposals like Goldman’s is that mindreading capacities are grounded in simulation; that is, simulative abilities come first and “whatever theoretical knowledge is achieved subsequently” (ibid. p. 225).

The opponent to the simulation theory is known as theory-theory. Theory-theory holds, roughly, that when we mindread, we access and utilize a theory of human behavior represented in our brains. It posits a theory of human behavior commonly known as ‘folk psychology’. Just like other folk theories, such as folk physics, it helps us to master our daily lives successfully. On this view, mindreading is essentially an exercise in theoretical reasoning.<sup>19</sup> When we predict behavior, for example, we utilize folk psychology in order to reason from representations of the target’s past and present circumstances and behavior (including verbal behavior), to representations of the target’s future behavior. For theory-theory, if there is just one mechanism, then metacognition depends on mindreading. Metacognition is merely the result of turning our mindreading capacities upon ourselves. In metacognition we just self-interpret ourselves.<sup>20</sup>

It is worth stressing at this point that defenders of the priority of mindreading faculty do not have to deny that we can access certain mental states. For example, Carruthers (2011) maintains that the mindreading faculty should access our sensory states:

It is obvious why the mindreading system should have access to vision and audition. For to interpret the intentions behind a smile, or a gesture, or a spoken phrase, the mindreading system would need access to perceptual representations that encode that smile, or gesture, or phrase... Everyone should predict, therefore, that people have transparent, non-interpretive, access to their own perceptual and imagistic states. (ibid. p. 51)

One might claim that this observation suggests the priority of metacognition over mindreading, but this suggestion arises due to the ambiguity of the meaning of the statement “access to our sensory states” as Byrne (2012) has pointed out. Consider a perception of someone’s smile. Access to this state might mean access to the information that the state carries—that someone is smiling—or to the information about one’s perceptual state—

<sup>19</sup> It should be remarked that what is relevant here is whether attribution of mental states is theory-like. Carruthers (2011) distinguishes his “modular” ISA theory from what he calls “*theorizing theories*”, which would be committed to a theorizing account of the developmental process. This refinement is irrelevant for my current purposes.

<sup>20</sup> For different interpretations of the view that mindreading is prior to metacognition see Gazzaniga (1995); Gazzaniga (2000); Gopnik (1993); Gopnik et al. (2004); Wilson (2002). This is the view endorsed by Carruthers himself. More precisely, in Carruthers (2000), where he presents his theory of phenomenal consciousness, he suggests that mindreading and metacognition are a unique mechanism with two different modes of access, one for perception (mindreading) and one for introspection (metacognition). In Carruthers (2006) he gives up this view and defends the priority of mindreading. Carruthers (2011) presents and carefully defends the more radical view that “our mode of access to the non-sensory aspects of our own minds is no different in principle from our access to the mental states of other people”. (p. 1)



that one sees someone smiling. It is clear that the sense in which uncontroversially mindreading requires access to one's perceptual state is the first one: mindreading should have access to the content of our perceptual states. Metacognition entails in turn the second sense, which involves representations to the effect that one is in certain state.

According to HOR theories, consciousness depends on metacognition. Now, if phenomenal consciousness is a necessary condition for mindreading then the thesis that mindreading is prior to metacognition is not compatible with HOR theories of consciousness. If the intuition presented in the previous section is correct, the priority of mindreading is not an option for HOR theories.

The only alternative available to HOR theories is that metacognition is prior to mindreading, endorsing some kind or other of simulation theory. The problem in this case is that this would, in turn, require that there had been an evolutionary pressure for metacognition; namely, that metacognition confers an adaptive advantage to an organism. Whereas there seems to be clear evolutionary advantage in mindreading for social interaction—independently on whether one thinks that the primary selection pressure has to do with enhanced capacities for cooperation (Tomasello 2008) or with “Machiavellian intelligence” (Byrne and Whiten 1988)—it is unclear what adaptive advantage metacognition would provide.

For instance, the claim that higher-order representation metacognition lacks biological function seems to be endorsed by David Rosenthal, one of the main proponents of HOR theories (Rosenthal 1997; Rosenthal 2005). Rosenthal (2005, 2008, 2012) maintains that phenomenal conscious states lack biological function, because most mental states seem to admit unconscious versions with similar causal powers, and that a theory like his, which predicts that phenomenal consciousness lacks biological function, is to be preferred. Rosenthal (2008) offers an alternative genesis of phenomenal consciousness, in cases such as beliefs and desires, without ascribing any biological function to these states. Rosenthal appeals, however, in his explanation to our mindreading capacity and his view faces, therefore, my objection: phenomenal consciousness seems to be prior to our mindreading ability. Furthermore, Rosenthal (2012) has also suggested that higher-order representation is a different psychological mechanism from metacognition. Two things should be remarked in reply. First, as noted in footnote 3, Rosenthal provides evidence for the claim that not all metacognition is a higher-order representation of the kind involved, according to the theory, in consciousness; but a reply to my argument would require that such higher-order representation does not depend on metacognition. This is a weaker claim and Rosenthal does not offer any argument against it. Second, if higher-order representation depends on metacognition we can make sense of the claim that consciousness lacks biological function, for it can be a by-product of metacognition—of course, only insofar as we could provide a biological function for metacognition. But, if it does not depend on metacognition and lacks biological function, then, as Aiello and Wheeler argue, it is implausible that the new neural wirings that higher-order representations would require are constructed and maintained.<sup>21</sup>

<sup>21</sup> Relatedly, one might suggest that metacognition is based on theory-theory and insist that metacognition is prior to mindreading by arguing that we first theorize about our own mental state and then about those of others. However, this alternative is, I think, untenable. The advantages of being able to theorize about the mental states of others are overwhelming—Machiavellian reasoning for example—in comparison with those, if any, of theorizing about our own mental states. Besides, it seems that we

Not all HOR theorists agree with Rosenthal. Gennaro (2004), for example, makes an interesting suggestion in the opposite direction in reply to Carruthers—who wonders in objection to actualist HOR theories: “What would have been the evolutionary pressure leading us to generate, routinely, a vast array of [actual] HOTs concerning the contents of our conscious experience?” (Carruthers (2000, 255) as quoted by Gennaro (2004, 51))—that might be recovered in the present debate. Gennaro, following Rolls (1999), proposes that “having actual HOTs allows for the correction of plans that result from first-order processing.” (2004, 52)<sup>22</sup> In his own proposal, Rolls (2004) submits that “part of the evolutionary adaptive significance of this type of higher order thought is that it enables correction of errors made in first order linguistic or in non-linguistic processing. Indeed, the ability to reflect on previous events is extremely important for learning from them, including setting up new long-term semantic structures”. Shallice (1988), for example, has also suggested that metacognition has evolved to supervise first-order, cognitive processes.

In reply, one should first note that it is not clear that such a role requires capacities beyond the mindreading ones. To understand why, it is crucial to distinguish metacognition in the intended sense—as cognition about (representing) one’s own cognition—from other uses in cognitive sciences. The term ‘metacognition’ is sometimes used to refer to any process above regular cognitive processes. In this sense, any process that makes use, or monitors, the output of a cognitive process would count as metacognitive. In the intended sense, however, the relation between the metacognitive process and the cognitive one is intentional or representational and not merely causal. With this distinction in hand, Carruthers (2009, sec. 5.1) has argued that alleged cases of ‘metacognition’ for executive monitoring and control are not cases of metacognition in the intended sense—where higher-order representation is required. Furthermore, if metacognition had evolved for monitoring and control we would expect it to be able to cognitively intervene in order to improve our learning, because we would expect to be able to monitor and control the progress of learning. However, there is no empirical evidence to that effect, quite the opposite—see Carruthers (2011, ch. 9 sec. 2) for a detailed review. As Carruthers concludes:

[S]tudies of the control that people exercise over their own learning show that it is indirect and behavior-based, and it seems that people lack the expected native capacities to control their own learning. Moreover, people’s judgments about their learning are equally indirect, and are based on a variety of heuristics and sensorily-accessible cues...At the very least we can conclude that there is no support for monitoring-for-control accounts of inner sense [the priority of metacognition] to be derived from the literature on human metacognition. (p. 272)

---

have far more behavioral evidence for others propensities than we have from ours. So, the claim that we theorize about our mental states by means of the same mechanism on which we theorize about those of others is plausible, but not so the other way around (for further discussion see Carruthers 2012). This lack of plausibility might explain why no one, to the best of my knowledge, has explored this line of reasoning in the debate about the relation between mindreading and metacognition.

<sup>22</sup> Gennaro proposes two further advantages of actualist theories. First, he notes that actual unconscious thoughts can more quickly become conscious resulting in introspective states, and even if this were wrong they might be thought as “a key stepping stone to the capacity for introspective consciousness”. It should be noted that both claims rest on the assumption that there is an adaptive advantage in introspecting our states—and not merely their content—and this is precisely what is at stake here.



It has also been proposed that an evolutionary advantage for a metacognition faculty might be reasoning about reasoning. For example, Rolls (2004) states that “another part of the adaptive value of a higher order thought system may be that by thinking about its own thoughts in a given situation, it may be able to better understand the thoughts of another individual in a similar situation, and therefore predict that individual’s behavior better” (ibid. p. 150). This option is not very plausible in the light of the empirical research on reflective thinking. For, as Carruthers (2011, ch. 9 sec. 3) notes, if metacognition had evolved for purposes of cognitive control we would expect people to have good native capacities to control, troubleshoot, and improve their own reasoning process. But we seem to have very poor natural competence in evaluating and reasoning about our own reasoning (Bos et al. 2008; Moshman 2004; Pillow 2002; Weinstock et al. 2004). Furthermore there are proposals, like the one presented by Carruthers that show that there is no need to appeal to metacognition and that mindreading abilities suffice for explaining the empirical data on reflective thinking (ch. 9 sec. 3).

It might well be the case that metacognition had evolved for different purposes than those considered here but I don’t know of any alternative proposal to that of monitoring and control. It might also be the case that metacognition did not evolve at all—being the result of a mutation or a by-product of another selected for trait. But considering the costs required for building and maintaining new brain mechanisms (Aiello and Wheeler, 1995) this alternative is not very plausible, thereby leaving HOR theories in check.

#### 4. Conclusions

In this paper I have argued that phenomenal consciousness is a necessary condition for our mindreading ability. This observation jeopardizes theories that maintain that phenomenal consciousness is a by-product of our mindreading ability such as Carruthers’ (2000).

My objection might be extended to other HOR theories on the reasonable assumption that metacognition depends on mindreading. To put pressure on other HOR theories, I have argued that they cannot endorse the view that metacognition and mindreading are independent cognitive mechanisms. If, as I have argued, our capacity to attribute experiences to others does not come for free by the development of a mindreading faculty that is based merely on observation of behavioral propensities, then the tenability of HOR theories depends on the plausibility of a functional explanation of the evolution of metacognition. I have offered some reasons to doubt that such an explanation will be provided.

#### REFERENCES

- Aiello, L. & Wheeler, P. 1995. The expensive tissue hypothesis: the brain and the digestive system in human and primate evolution. *Current Anthropology* 36: 199-221.
- Armstrong, D. 1968. *A Materialist Theory of the Mind*. London: Routledge.
- Block, N. 2007. Wittgenstein and Qualia, *Philosophical Perspectives* 21(1): 73-115.
- Block, N. 2011. The Higher Order Approach to Consciousness is Defunct, *Analysis* 71 (3): 419-431.
- Bos, M.; Dijksterhuis, A. & van Baaren, R. 2008. On the goal-dependency of unconscious thought. *Journal of Experimental Social Psychology* 44: 1114-1120.

- Bugnyar, T. & Heinrich, B. 2006. Pilfering ravens, *corvus corax*, adjust their behavior to social context and identity of competitors. *Animal Cognition* 9: 369-376.
- Byrne, A. 2004. What Phenomenal Consciousness is Like. In Rocco J. Gennaro, ed. *Higher-Order Theories of Consciousness: An Anthology*. Amsterdam: John Benjamins Publishing Company.
- Byrne, A. 2012. Review of Peter Carruthers "The opacity of mind: An integrative theory of self-knowledge". *Notre Dame Philosophical Review* [online].
- Byrne, R. & Whiten, A. 1988. *Machiavellian Intelligence*. Oxford: Oxford University Press.
- Chalmers, D. 2003. The content and epistemology of phenomenal belief. In Quentin Smith and Aleksandar Jokic, eds. *Consciousness: New Philosophical Perspectives*. Oxford: Oxford University Press.
- Chalmers, D. 2003b. Consciousness and its place in nature, In Stephen P. Stich and Ted A. Warfield, eds. *Blackwell Guide to the Philosophy of Mind*. Malden, MA: Blackwell.
- Carruthers, P. 2000. *Phenomenal Consciousness: A Naturalistic Theory*, 1st edition. Cambridge: Cambridge University Press.
- Carruthers, P. 2006. *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Oxford: Oxford University Press.
- Carruthers, P. 2009. How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences* 32(2): 121-138.
- Carruthers, P. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.
- Carruthers, P. & Veillet, B. 2007. The phenomenal concept strategy. *Journal of Consciousness Studies* 14: 212-236.
- Gazzaniga, M. 1995. Consciousness and the cerebral hemispheres. In: Michael S. Gazzaniga, ed. *The Cognitive Neurosciences*. Cambridge, MA: MIT Press.
- Gazzaniga, M. 2000. Cerebral specialization and inter-hemispheric communication: does the corpus callosum enable the human condition? *Brain* 123: 1293-1326.
- Gennaro, R. 1996. *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. Amsterdam: John Benjamins Publishing Company.
- Gennaro, R. 2004. Higher-order thoughts, animal consciousness, and misrepresentation: A reply to Carruthers and Levine. In: Rocco Gennaro, R., ed. *Higher-Order Theories of Consciousness*, Amsterdam: John Benjamins Publishing Company.
- Gennaro, R. 2012 *The Consciousness Paradox: Consciousness, Concepts, and Higher-Order Thoughts*. Cambridge, MA: MIT Press.
- Goldman, A. & Shanton, K. 2010. The case for simulation. In: A. Leslie & T. German, eds. *Handbook of Theory of Mind*. Brighton: Taylor & Francis Group.
- Goldman, A.I. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, illustrated edition. Oxford: Oxford University Press.
- Gopnik, A. 1993. The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16: 1-14.
- Gopnik, A.; Glymour, C.; Sobel, D.; Schulz, L.; Kushnir, T. & Danks, D. 2004. A theory of causal learning in children: Causal maps and bayes nets. *Psychological Review* 111: 1-31.
- Gray, H.; Gray, K. & Wegner, D. 2007. Dimensions of mind perception. *Science* 315 (5812): 619.
- Har,e B. & Tomasello. M. 2005. Human-like social skills in dogs? *Trends in Cognitive Sciences* 9: 438-444.
- Jackson. F. 1982. Epiphenomenal Qualia. *Philosophical Quarterly* 32: 127-136.
- Kriegel. U. 2003. Consciousness, higher-order content, and the individuation of vehicles. *Synthese* 134 (3): 477-504.
- Kriegel, U. 2009. *Subjective Consciousness: A Self-representational Theory*. Oxford: Oxford University Press.
- Lycan, W.G. 1996. *Consciousness and Experience*. Cambridge, MA: MIT Press.
- Mandik. P. 2010. Swmap Mary's revenge: deviant phenomenal knowledge and physicalism, *Philosophical Studies* 148 (2): 231-247.
- Millikan, R.G. 1984. *Language, Thought and Other Biological Categories*. Cambridge, MA: MIT Press.

- Millikan, R.G. 1989. Biosemantics. *Journal of Philosophy* 86: 281-297.
- Moshman, D. 2004. From inference to reasoning: The construction of rationality. *Thinking and Reasoning* 10: 221-239.
- Nichols, S. & Stich, S.P. 2003 *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, illustrated edition. Oxford: Oxford University Press.
- Nida-Rumelin, M. 1996. Pseudonormal Vision: An actual case of Qualia Inversion? *Philosophical Studies*, 82: 145-157.
- Papineau, D. 1993. *Philosophical Naturalism*. Oxford: Blackwell.
- Papineau, D. 1999. Mind the gap. *Philosophical Perspectives*. Web version at: <http://www.kcl.ac.uk/schools/hums/philosophy/mindthegap.html>
- Peacocke, C. 1995. *A Study of Concepts*. Cambridge, MA: MIT Press.
- Pillow, B. 2002. Children's and adults' evaluation of certainty of deductive inference, inductive inference, and guesses. *Child Development* 73: 799-792.
- Rolls, E. 1999. *The brain and emotion*. Oxford: Oxford University Press.
- Rolls, E. 2004. A higher order syntactic thought (host) theory of consciousness. In: Rocco Gennaro, ed. *Higher-Order Theories of Consciousness*. Amsterdam: John Benjamins Publishing Company.
- Rosenthal, D. 1997. A theory of consciousness. In: Ned Block, Owen J. Flanagan & Guven Guzeldere, eds. *The Nature of Consciousness*. Cambridge, MA: MIT Press.
- Rosenthal, D. 2005. *Consciousness and Mind*. Oxford: Oxford University Press.
- Rosenthal, D. 2008. Consciousness and its function. *Neuropsychologia* 46(3): 829-840.
- Rosenthal, D. 2011. Exaggerated reports: reply to Block, *Analysis* 71 (3): 431-437.
- Rosenthal, D. 2012. Higher-order awareness, misrepresentation and function *Philosophical Transactions* 376: 1424-1438.
- Sebastian, M.A. 2012. Experiential Awareness. Do you prefer *It to Me* ? *Philosophical Topics* 40(2): 155-177.
- Shallice, T. 1988. *From neuropsychology to mental structure*. Cambridge: Cambridge University Press.
- Stulp, G.; Emery, N.; Verhulst, S. & Clayton, N. 2009. Western scrub-jays conceal auditory information when competitors can hear but cannot see. *Biology Letters* 5: 583-585.
- Stoljar, D. 2005. Physicalism and Phenomenal Concepts. *Mind & Language* 20 (5): 469-494.
- Strawson, P.F. 1959. *Individuals: An Essay in Descriptive Metaphysics*. London: Routledge.
- Tomasello, M. 2008. *Origins of Human Communication*. Cambridge, MA: MIT Press.
- Tye, M. 1999. Phenomenal Consciousness: The explanatory gap as a cognitive illusion. *Mind*, 108 (432): 705-725.
- Udell, M.; Dorey, N. & Wynne, C. 2008. Wolves outperform dogs in following human social cues. *Animal Behavior* 76: 1767-1773.
- Weinstock, M.; Neuman, Y. & Tabak, I. 2004. Missing the point or missing the norms? Epistemological norms as predictors of students' ability to identify fallacious arguments. *Contemporary Educational Psychology* 29: 77-94.
- Weisberg, J. 2011. Abusing the notion of what-it's-like-ness: A response to Block, *Analysis* 71 (3): 438-443.
- Wilson, T. 2002. *Strangers to Ourselves*. Cambridge, MA: Harvard University Press.
- Zahavi, D. 2005. *Subjectivity and Selfhood: Investigating the first-person perspective*. Cambridge, MA: MIT Press.
- Zahavi, D. 2006. Two takes on a one-level account of consciousness. *Psyche* 12 (2) [online].

**MIGUEL ÁNGEL SEBASTIÁN** is Associate Research Fellow at the Instituto de Investigaciones Filosóficas at Universidad Nacional Autónoma de México. His research focuses on the philosophy of mind and cognitive sciences.

**ADDRESS:** Instituto de Investigaciones Filosóficas, UNAM. Circuito Mtro Mario de la Cueva. Ciudad Universitaria Del. Coyoacán. México D.F. 04510. email: [msebastian@gmail.com](mailto:msebastian@gmail.com). URL: <http://mindingthebrain.wordpress.com/>