

Consciousness and Theory of Mind: a Common Theory?

Keywords: Consciousness, Higher-Order Theories, Theory of Mind, Mind-reading, Metacognition.

1 Introduction

Many of our mental states are phenomenally conscious. It *feels* a certain way, or to borrow Nagel's expression, *there is something it is like* to be in these states. Examples of phenomenally conscious states are those one 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. Higher-Order Representational (HOR) theories¹ attempt to provide such an explanation. 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. According to these theories, 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. 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 mental states thanks to the mindreading ability. We attribute beliefs, perceptions, feelings or desires to other people and predict and explain their behaviour 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 argue against HOR theories of consciousness by showing that the claim that phenomenal consciousness depends on a ToM is not plausible.

¹ See, for instance, Armstrong (1968); Carruthers (2000); Gennaro (1996); Lycan (1996); Rosenthal (1997, 2005)

Most HOR theories maintain that phenomenal consciousness depends on metacognition. Some philosophers, however, have maintained that it depends on our mindreading capacities. In section 2 I will argue against this last 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 is a plausible option for HOR theories.

2 Phenomenal Consciousness does not Depend on Mindreading.

HOR theories of consciousness try to explain the difference between phenomenally conscious states and other kinds of states in terms of some relation obtaining between the phenomenally conscious state and some sort 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); Rosenthal (1997, 2005)) the latter Higher-Order Perception (HOP) or 'inner-sense' theories (Amstrong (1968); Carruthers (2000); Lycan (1996)). According to the former theories, when I have 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 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 (Rosenthal (1997, 2005)) or by the disposition to raise a higher-order representation (Carruthers (2000)). Carruthers' dispositionalist theory is the initial target of this paper.

According to HOR theories, beings lacking metacognition lack thereby phenomenal consciousness. Carruthers (2000) further claims that the ability of mindreading is required.²

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 Theory of Mind faculty combined with the truth of some version of 'consumer semantics'

² Some philosophers consider this to be a reason for rejecting these theories. It is too demanding, for it requires precisely a Theory of Mind and most animals and arguably human babies lack it. 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 have phenomenally 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, but surely, when comparing alternative theories, one that doesn't have this consequence is to be preferred.

(Millikan (1984, 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 organism which 'consumes' that state; for instance, what a state represents will depend on the kinds of inferences which the cognitive system is prepared to make in the presence of that state.

Certain mental states are recognized as mental representations by the Theory of Mind and it is in virtue of their availability to ToM faculty, as a consumer system, that the perceptual states in question acquire a dual content. These states are phenomenally conscious states:

Each phenomenally conscious experience has its distinctive form of subjectivity by virtue of acquiring a higher-order analogue content which precisely mirrors, and represents as subjective, its first-order content. (ibid. p. 243)

Our evolutionary ancestors would have had first-order representational concepts for many features of the environment (red, green, etc); then the development of a ToM 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 ToM 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 ToM.

This new content, *seems red*, is a by-product of a mindreading faculty, which builds up the *is/seems* distinction. Besides there being a first order representation of redness, there is also second-order representation of seeming-redness.

Carruthers introduces an interesting proposal defending higher-order representational 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.

My purpose in this objection is to show the implausibility of a theory of consciousness according to which having a phenomenally conscious experience

depends on our mindreading abilities.³ 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.

My opponent would argue that conscious experiences are not necessary for developing a ToM along these lines: creatures can see objects in the environment and the response of other organisms to those objects and their properties. 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. Similarly, my opponent would argue, in the case of experiences: when people attribute to others sensory states there is no reason for attributing feeling, we just attribute to them states that track certain properties.

That seems to me to be completely misguided as it is dramatically clear in the case of pains or orgasms. The kind of mental state ascription mentioned above is very different from the kind of attribution we usually do. How can one ascribe others with mental states that *feel* in a certain way if one has never been in a mental state that *feels*? It seems to me that the kind of attribution would be completely different in this case.⁴ For illustration, consider Sally who has never had an orgasm in her life. Sally knows that she has never had an orgasm. She can nevertheless ascribe orgasms to other people. As a matter of fact, she is really good in that task and she can always recognize when her partners are having an orgasm or just faking given their behavioral response. Surely the kind of mental state Sally attributes to her partners or, for instance, actors when seeing a film, is a phenomenally conscious mental state. My intuition is that clearly, after she has an orgasm for the first time, the kind of experience that 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 attributes to others a similar sensation when they are having one.

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. 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. So, phenomenal consciousness cannot be a by-product of our mindreading capacities, precisely because our mindreading capacities require phenomenally conscious

³ Carruthers (2009) seems to take *feelings* as inputs for a mindreading ability.

⁴ This is independent of whether my ascription of mental states to myself or others is due to a simulation theory or purely theoretical.

mental states.

One possible alternative theory, not clearly in the spirit of Carruthers' one, would maintain that our theory of mind evolved in two steps. In a first step a proto-theory of mind attributes states with certain functional role. A mental state is phenomenally conscious in virtue of being available to this proto-theory of mind. In a second step a *full-blown* ToM evolves and allows the attributions of phenomenally conscious states to others.

The problem of this reply is that, according to Carruthers, the functional role attributed by the proto-theory of mind exhausts the phenomenal character and the proto-theory of mind already allows the attribution of mental states with this role. So, there is no evolutionary advantage in attributing phenomenally conscious mental states and therefore there is no justification for the evolution of the mechanisms underlying this new *full-blown* ToM.

If my intuition 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.⁵ 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.

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 reason is that i) the previous intuition suggests that phenomenal consciousness is a necessary condition for our mindreading ability and ii) according to higher-order 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 Nichols and Stich proposal, 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

⁵ See Carruthers (2009) for an excellent review.

mindreading depends on the mechanisms that evolved for metacognition) or the other way around.

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 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). Goldman and Shanton (2010)

The opponent to the simulation theory is known as theory-theory. Theory-theory holds 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. 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.⁶

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. Theory-theory is not an option for HOR theories, if the intuition presented in the previous section is correct.

The only alternative available to HOR theories is that metacognition is prior to mindreading, endorsing some kind or other of simulation theory. This would,

⁶ For different interpretations of the view that mindreading is prior to metacognition see Gazzaniga (1995, 2000); Gopnik (1993); 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, in favor of the one presented here.

however, 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 (the so called 'machievellian intelligence' is clearly beneficial in social environments) I know of no HOR theory that makes such a claim and it is hard to see what such an advantage would be.

For instance, the claim that metacognition lacks biological function seems to be endorsed by David Rosenthal, one of the main proponents of HOR theories (1997; 2005). Rosenthal maintain that phenomenal conscious states lack biological function, because most mental states seem to admit unconscious versions with similar causal powers and a theory like his that 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 appealing to any notion of function. Rosenthal appeals, however, in his explanation to the mindreading capacity and his view faces, therefore, my objection: phenomenal consciousness seems to be prior to our mindreading ability.

It has been suggested that metacognition has evolved to supervise first-order, cognitive processes (Shallice (1988)), but 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 is intentional or representational and not merely causal. With this distinction in hand, it can be shown, as Carruthers (2009, sec. 5.1) does, that alledged cases of 'metacognition' for executive monitoring and control are not cases of metacognition in the intended sense where higher-order representation is required. This leaves the thesis that metacognition is prior to mindreading unsupported.

4 Conclusion

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 press on other HOR theories, I have argued that HOR theories cannot endorse the view that metacognition and mindreading are independent cognitive mechanisms.

The tenability of HOR theories depends, therefore, 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

- Armstrong D (1968) *A Materialist Theory of the Mind*. London: Routledge
- Carruthers P (2000) *Phenomenal Consciousness: A Naturalistic Theory*. Cambridge University Press
- Carruthers P (2006) *The Architecture of the Mind: massive modularity and the flexibility of thought*. 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
- Gazzaniga M (1995) Consciousness and the cerebral hemispheres. In: Gazzaniga M (ed) *The Cognitive Neurosciences*, MIT Press
- Gazzaniga M (2000) Cerebral specialization and inter-hemispheric communication: does the corpus callosum enable the human condition? *Brain* 123:1293–1326
- Gennaro RJ (1996) *Consciousness and Self-Consciousness: A Defense of the Higher-Order Thought Theory of Consciousness*. John Benjamins
- Goldman A, Shanton K (2010) The case for simulation. In: Leslie A, German T (eds) *Handbook of Theory of Mind*, Psychology Press
- Goldman AI (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, illustrated edition edn. Oxford University Press, USA
- Gopnik A (1993) The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences* 16:1–14
- 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(July ,):281–97
- Nichols S, Stich SP (2003) *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, illustrated edition edn. Oxford University Press, USA
- Papineau D (1993) *Philosophical Naturalism*. Blackwell
- Peacocke C (1995) *A Study of Concepts*. The MIT Press

-
- Rosenthal D (2008) Consciousness and its function. *Neuropsychologia* 46(3):829–840
- 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
- Shallice T (1988) *From neuropsychology to mental structure*. Cambridge University Press
- Wilson T (2002) *Strangers to Ourselves*. Harvard University Press