

**Comments on “Evolving Self-Consciousness” by Peter Carruthers, Logan Fletcher,
and J. Brendan Ritchie**

JeeLoo Liu
California State University, Fullerton

This article examines the evolutionary origins of the form of self-consciousness involved in our knowing “what it is that we see, hear, feel, judge, want, or decide.” This is more an issue of self-knowledge, and in particular, self-knowledge of the current content of our perception, sensation, desire and thought in general. Carruthers *et al* discuss self-consciousness as rooted in a form of adaptation, which is an alteration of the organism’s structure or function as a result of natural selection, so that the organism can be better fitted to survive in its environment. The question for our interest is: What evolutionary benefits are provided by our having the capacity of knowing what we are thinking or feeling? Or we may ask: How did humans’ capacity for knowing *that* they think and *what* they think evolve as the result of humans’ adaptation to the environment?

Two proposals on the adaptive basis of self-consciousness are contrasted and evaluated against empirical evidence. The first theory takes the first-person approach and explains the emergence of self-consciousness as “an adaptation designed initially for metacognitive monitoring and control.” As an adaptation in human beings, self-consciousness enables humans to “become more efficient and reliable cognizers, and can make better and more adaptive decisions as a result.” By this account, the knowledge of our current thought is needed for our being better thinkers and more efficient actors. Therefore, self-consciousness emerged as an independent adaptation. Some proponents of this theory treat the two capacities as independent of each other, while some further argue that capacities for mindreading could be dependent on capacities for self-consciousness. In this paper, the target theory is merely the independence claim, which the authors call the “minimalist” first-person view.

The second theory takes the third-person approach and locates the evolutionary origin of self-consciousness in “the prior evolution of a mindreading system,” which is primarily used on others but can be directed toward the self as in the cases of self-consciousness. By this account, humans’ self-consciousness is based on humans’ other capacities for mindreading and it evolved primarily for social purposes. Self-consciousness is not an independent capacity evolved as a result of need. It is rather a derived faculty.

The bottom line in the debate between the two theories is then not just about the evolutionary basis of self-knowledge, but also about the evolutionary necessity of a separate faculty to introspect. Carruthers *et al* support the second theory. In this article, they aim to show that self-consciousness is a derived ability from the adaptation of a mindreading system initially intended for social purposes. Hence, self-consciousness is not itself an adaptation.

Humans do possess both capacities for mindreading and capacities for self-knowledge, but which one emerged first as an adaptation? What kind of empirical evidence could convince us to opt for one theory or the other? Carruthers *et al* argue that if self-consciousness were

indeed a separate adaptation, then it must meet the following three criteria, which they call “adaptive signatures”:

- (1) If a capacity is itself an adaptation, then it should enable humans “to do well what it was allegedly selected *for*.” Evidence of good performance should be species-wide rather than idiosyncratic.
- (2) If a capacity is an adaptation for a particular function, then it should emerge as early as when that function is needed for the individual’s development.
- (3) If a capacity is a universal adaptation for human beings, then it should emerge steadily in individuals across environmental variations.

With these three criteria, they compare the two accounts of the evolutionary basis for self-consciousness. To begin with, humans’ metacognitive abilities and our capacities for self-monitoring or self-control do not seem to be universally successful, they emerge comparatively late in childhood, and there are wide individual differences resulting from different learning history or cultural training. This shows that humans do not have a native competence for metacognitive capacities. Therefore, our metacognition does not seem to be an adaptation in itself.

On the other hand, mindreading seems to have passed the three criteria. Humans have demonstrated success at mindreading, as can be observed in our cooperative or competitive activities. It thus meets criterion 1. Carruthers *et al* also cite evidence in developmental psychology that the core of humans’ mindreading system is up and running at age 2. They thus argue that mindreading as adaptation meets criterion 2 above, namely, that mindreading capacity emerged as early as it is needed. Further, the hypothesis meets criterion 3, as evidence shows that normal human children acquire this capacity within similar time frame in spite of environmental variances. This would support the view that mindreading is an independent adaptation.

This is the first step of their argument. The second step of their argument is to further support the derivability of self-consciousness from mindreading capacities. Since the third-person approach takes mindreading as primary, and self-consciousness as derivative from mindreading, the empirical evidence needed in individual development is that the capacity of self-consciousness cannot be present without the mindreading capacity. In other words, we should find no creatures that are capable of self-knowledge but are incapable of mindreading. This becomes what the authors take to be the test case for the verity of the third-person hypothesis.

According to the received view, mindreading is not an all or none capacity, but a capacity evolved in degrees. There are two stages of mindreading: Stage 1 involves understanding others’ goals, perceptual access to the world, and states of knowledge and ignorance. Stage 2 mindreading involves further understanding of the beliefs and false beliefs of others, as well as the ways in which agents can be misled by appearances.

In consultation with data from comparative psychology, Carruthers *et al* argue that studies showing other primates to be exhibiting both state 1 mindreading capacity and some awareness of their own desires, perception, and state of knowledge, etc. would not undermine

the third-person theory, as the data would be neutral as to whether self-consciousness is a derived capacity. Some other studies, however, show that other primates have failed stage 2 mindreading tests, but have exhibited apparent self-awareness such as in their (i) having the sense of uncertainty and (ii) being able to detect misleading appearances. These phenomena seem to show that the primates in question have the sense of their own judgment or perception (and even have doubts about it), and yet they could not successfully discern others' false beliefs or misleading appearances. If so, then it would be a case where creatures have self-consciousness without the accompanying stage 2 mindreading capacities.

Carruthers *et al* introduced alternative interpretations of the phenomena presented by another paper by Carruthers. Under these different explanations, these animals do not really possess metacognitive capacity. They merely rely on prior perception of the size of object and subsequent belief. So what the animals demonstrate is not conceptual repertoire that involves self-consciousness, but some physical beliefs about the size of the object. Hence there is no compelling case of any creature's having self-consciousness without mindreading capacities.

The advantage of the third-person approach is its theoretical economy – it avoids positing an extra mental faculty especially evolved for introspective knowledge. However, the first-person approach is intuitively appealing: from introspection, we do seem to know what we think, how we feel and what we desire even if our self-knowledge is not always veritable or even reliable. Lacking expertise on evolutionary psychology, I cannot decide whether the first-person approach or the third-person approach to the evolutionary origin of self-consciousness is the right theory. I shall only examine the two arguments presented in this paper.

To being with, I don't find the argument from comparative psychology convincing or even relevant. If data from animal psychology can be interpreted with different outcomes, then what the data really show is questionable. Do we have a case of creatures with self-consciousness without mindreading capacities in apes or other primates? It seems that both the faculty of self-consciousness and mindreading would have to be greatly qualified to apply to other primates. Even if a case could be established for apes' having self-awareness without stage 2 mindreading capacities, it does not settle the case for the evolutionary basis of humans' self-consciousness, which is heavily language-based. Since this argument is merely used as a defensive strategy against the interpretation that proponents of the first-person approach employ to defeat the third-person approach, I shall not spend more time on this argument.

The first argument links the emergence of mental traits in child development with the evolutionary history of these mental traits. I question whether the three criteria they set as "adaptive signatures" can really provide empirical testability of either approach. The authors' backward inference into the history of the evolution of mind is based on observation of current child development. However, current child development is already heavily impacted by cultural factors, as Carruthers *et al* acknowledge that children in the modern world are "primed for credulity." The third criterion they use, namely, that the capacity should emerge steadily in spite of environmental variances, can only be accepted if the adaptation of our mental trait is merely biologically based. Once we allow the possibility of evolution by culture, especially in view of the fact that many of our mental capacities are based on language acquisition, cultural differences should naturally bring about different performances or developments in individuals. What we observe in child development is not

necessarily the indicator for the evolutionary process of human mind. Thus I reject their criterion 3.

Furthermore, the first argument includes some sweeping claims about the performance rate of humans' mindreading and self-consciousness capacities. I want to challenge the following pair of claims:

1. "Mindreading abilities are robust and early to emerge across individuals and cultures in the absence of instruction, and issue in highly successful performance."
2. "People are only modestly successful, at best, in controlling their own cognitive processes effectively, and many either employ strategies that are actually maladaptive, or make no attempt to control their cognitive processes at all even when it would be adaptive to do so."

If mindreading capacity could be developed in stages, then so could metacognitive capacities. I shall propose the following three-stage theory for self-consciousness:

1. Stage-1 self-consciousness: knowing one's goals, perceptual access to the world, and states of knowledge and ignorance.
2. Stage-2 self-consciousness: understanding of one's beliefs and false beliefs, as well as the ways in which one can be misled by appearances.
3. Stage-3 self-consciousness: having self-monitoring and self-control of one's cognitive processes as well as actions, in the way that is advantageous for survival. [This kind of self-consciousness would be a form of "meta-reasoning" or some kind of reflective reasoning which can override one's swift, intuitive judgments.]

If we separate these stages of self-consciousness alongside stages of the mindreading system, then it is not clear that the mindreading capacities evolved earlier in history or that they developed sooner in children. Young children probably have stage-1 self-awareness before they develop stage-1 mindreading. The case about Maxi in the "false belief test" can be interpreted to show that children before the age of four interpret Maxi's belief by attributing to him what they themselves believe. They attribute their own belief to Maxi, without realizing that Maxi cannot have the same belief (being ignorant of the location switch). So they know their mind better than they know the minds of others. As for stage-2 mental capacities, I suspect that people do not fare better in their understanding of others' false beliefs and misrepresentation than they do with their own. Stage-3 metacognition involves self-monitoring and self-control, and it can be granted that people do not perform well in this respect. But they perform much worse when it comes to stage-3 mindreading, if there is such a stage. In other words, when we compare self-consciousness and mindreading at comparable stages, we do not find that people generally are good at or are more successful in the latter than in the former.

Finally, let us go back to the initial question that interests us: What adaptive function does self-consciousness serve if any? I venture to offer a "just-so theory": our self-reflective metacognitive abilities enable us to slow down in our responses to environmental stimuli, so

that we do not act on impulse as most animals do. Our reflective mental trait would be evolutionarily beneficial for us since we can avoid rash behavior that brings peril to ourselves. Our self-interpretation and self-attribution aim to fit the way we view ourselves as well as the way we view the situations in which we find ourselves. Granted, whether individuals are *good at* judging the content of their current mental states or at monitoring and controlling themselves would depend on individual personality, cultural training and other contextual factors. As (Fletcher & Carruthers 2012) points out, the capacity of “reflective reasoning” is “to a large extent an acquired *habit*, which has been cultivated more successfully in some individuals than in others.” But these individual variances of stage-3 metacognition should not be taken to indicate that self-consciousness itself is not a *universal* human trait or that it is not *robust* enough. The function of our self-consciousness in these metacognitive activities is not to protect us from making *false* judgments, but to prevent us from making *rash* judgments. I believe that this is an important trait of human beings and thus I am inclined to accept the first-person approach to view self-consciousness as an independent adaptation.