involve the direct apprehension of verbal, imaginal, or other symbols and does not involve sensory awareness as DES defines that term. The apprehension of an unsymbolized thought may involve the apprehension of some sensory bits, so long as those sensory bits are not organized into a coherent, central, thematized sensory awareness. Thus, I believe that unsymbolized thinking is a perceptual event, just as are inner speech, visual imagery, and feelings; it is therefore not purely propositional and is therefore not a threat to the mindreading-is-prior view.

Access to propositional attitudes is interpretative. Far from being neutral, DES lends empirical support to the main thrust of Carruthers' analysis that propositional attitudes are interpreted, not observed. The DES procedure trains subjects carefully, repeatedly, and iteratively (Hurlburt & Akhter 2006; Hurlburt & Heavey 2006; Hurlburt & Schwitzgebel 2007) to distinguish between directly observed (Carruthers' "perceptual") events and all else; that training typically requires several days. DES tries, moment by moment, to cleave to the directly observed and to bracket all that is inferred, supposed, presupposed. There is no a priori assumption about what is or is not directly observable. Attitudes are not singled out; if an attitude is directly observed at the moment of some beep, then that attitude is the proper target of DES. If not, then it isn't.

As a result of 30 years of carefully questioning subjects about their momentary experiences, my sense is that trained DES subjects who wear a beeper and inspect what is directly before the footlights of consciousness at the moment of the beeps almost never directly apprehend an attitude. Inadequately trained subjects, particularly on their first sampling day, occasionally *report* that they are experiencing some attitude. But when those reports are scrutinized in the usual DES way, querying carefully about any perceptual aspects, those subjects retreat from the attitude-was-directly-observed position, apparently coming to recognize that their attitude had been merely "background" or "context." That seems entirely consonant with the view that these subjects had initially inferred their own attitudes in the same way they infer the attitudes of others. (I note that subjects do not similarly retreat from their initial reports about unsymbolized thinking; they continue to maintain that the unsymbolized thought had been directly observed.)

What monkeys can tell us about metacognition and mindreading

doi:10.1017/S0140525X09000685

Nate Kornell,[a] Bennett L. Schwartz,[b] and Lisa K. Son[c]

[a]*Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095-1563;* [b]*Department of Psychology, Florida International University, Miami, FL 33199;* [c]*Department of Psychology, Barnard College, New York, NY 10027.*

nkornell@ucla.edu **http://nkornell.bol.ucla.edu/**
bennett.schwartz@fiu.edu **www.fiu.edu/~schwartb**
lson@barnard.edu **http://lisason.synthasite.com/index.php**

Abstract: Thinkers in related fields such as philosophy, psychology, and education define metacognition in a variety of different ways. Based on an emerging standard definition in psychology, we present evidence for metacognition in animals, and argue that mindreading and metacognition are largely orthogonal.

The target article proposes that "mindreading is prior to metacognition," meaning that just as we know the minds of others by observing what they do, we know our own minds by observing what we do. According to this view, metacognition – that is, cognition about one's own cognition – requires mindreading abilities. Rhesus monkeys (*Macaca mulatta*) do not appear to possess mindreading abilities (Anderson et al. 1996; but see Santos et al. 2006). Here we present evidence, however, that rhesus monkeys are metacognitive. We offer a different definition of mindreading than that used by Carruthers, and we contend that the mechanisms of mindreading and metacognition are largely orthogonal.

The target article reports in detail on only a few seminal studies of metacognition in animals (see Smith et al. 2003; Smith & Washburn 2005). We begin by elaborating on subsequent studies that provide evidence of animal metacognition (reviewed by Kornell, in press). For example, Hampton (2001) tested monkeys in a modified delayed match-to-sample task: On each trial, a sample picture was presented on a touch-sensitive computer monitor, and then, after a delay, the same sample picture was presented among three distractors, and the subject had to touch the sample. On some trials, after viewing the sample, the monkey could choose to skip the test and receive a small reward. If the monkey instead chose to take the test, he could earn a large reward, or, if his response was incorrect, forfeit reward completely. Memory accuracy was better on self-selected test trials than on mandatory test trials. It appears that the monkeys chose to take the test when they knew that they knew the answer, in the same way that a student raises her hand in class when she knows that she knows the answer (see Suda-King 2008, for similar results in orangutans).

In another study, two male rhesus monkeys were asked, essentially, to bet on their memories (Kornell et al. 2007). A given monkey was shown six pictures sequentially for "study," followed by a display of nine pictures presented simultaneously, one of which had been "studied." The monkey's task was to select the studied picture. After he responded, two "risk" icons were presented, which allowed the monkey to bet his tokens (which could be exchanged for food). A high-risk bet resulted in the gain of three tokens if the monkey had responded correctly, but a loss of three tokens otherwise. Choosing low-risk resulted in a sure gain of one token. The monkeys made accurate confidence judgments: They bet more after correct responses than after incorrect responses. This finding was especially impressive because the monkeys were originally trained on tasks that involved neither pictures nor remembering (e.g., select the longest line); following that training, they were able to respond metacognitively beginning on the first day of the picture-memory task. The monkeys appear to have learned a general metacognitive response, not one that was task-specific.

In addition to being able to make judgments about their memories, monkeys have demonstrated that they can choose behaviors, based on metacognition, that advance their own knowledge – that is, they have demonstrated metacognitive control (see Nelson & Narens 1990). To investigate this ability, we allowed two monkeys to request information when they were uncertain, just as a person might ask for a hint when answering a difficult question (Kornell et al. 2007). The monkeys could request a "hint," that is, a blinking border that surrounded the correct response, on some trials in a list-learning experiment. As the monkeys' response accuracy on no-hint trials improved steadily, their rate of hint requests showed a corresponding decline. By requesting hints when they were unsure, the monkeys went beyond making an uncertain response; they took steps to rectify their ignorance.

Based on the studies described above, we conclude that monkeys have metacognitive abilities – that is, they can monitor the strength of their own internal memory representations. According to the target article, these findings fall short of metacognition, however. Carruthers writes, "It is only if a human reports that she acted as she did, not just because she *was* uncertain, but because she was *aware of being* uncertain, that there will be any conflict [with the metacognition is prior account]" (sect. 5.2, para. 3). We do not agree that metacognition requires awareness; we have previously argued that the metacognitive abilities that animals possess are not necessarily conscious (Kornell, in press; Son & Kornell 2005; also see Reder 1996).

For example, a monkey might make a high-risk bet without being aware that it is monitoring its internal memory trace.

We are not arguing that mindreading cannot subsume metacognitive functions. Indeed, we can learn much about ourselves by observing our own behavior: for example, after playing a round of golf, we decide we are not quite ready for the pro tour. Moreover, numerous experiments have shown that metacognition is largely based on unconscious inferential processes, not direct examination of memories; for example, we infer that we know something well based on the fluency (i.e., ease and speed) with which it comes to mind (Schwartz et al. 1997).

Given the way we, and many other cognitive psychologists, define metacognition, we assert that it is likely that metacognition and mindreading are separate processes. The argument that one should only see metacognition in species that can mindread is, to the best available evidence, false. For example, some have suggested that dogs, which have shown no metacognitive abilities but show high levels of social cognition, may have rudimentary mindreading abilities (Horowitz, in press; Tomasello et al. 1999). Conversely, we offer rhesus monkeys as a case study in a metacognitively competent animal that fares poorly at mindreading. In the tasks we describe, metacognitive processing can lead to positive outcomes that are evolutionarily adaptive. Indeed, metacognitive monitoring seems to have its own rewards.

Metacognition without introspection

doi:10.1017/S0140525X09000697

Peter Langland-Hassan

Department of Philosophy, The Graduate Center of the City University of New York, New York, NY 10016.

PLangland-Hassan@gc.cuny.edu

https://wfs.gc.cuny.edu/PLangland-Hassan

Abstract: While Carruthers denies that humans have introspective access to cognitive attitudes such as belief, he allows introspective access to perceptual and quasi-perceptual mental states. Yet, despite his own reservations, the basic architecture he describes for third-person mindreading can accommodate first-person mindreading without need to posit a distinct "introspective" mode of access to *any* of one's own mental states.

Carruthers argues that passivity symptoms (e.g., thought insertion) in schizophrenia result not from a special metacognitive deficit, but from "faulty data being presented to the mindreading system" (sect. 9, para. 2). Although I endorse Carruthers' Frith-inspired (Frith et al. 2000a; 2000b) appeal to efference-copy deficits in the explanation of passivity symptoms, his claim that the mindreading faculty itself is undamaged raises questions. First, any attribution of one's own thoughts to another is equally a mistake in first- and third-person mindreading (false positives count as errors just as much as false negatives do). Carruthers should therefore hold that mindreading – first- *and* third-person – is deficient in these forms of schizophrenia; this still allows him to deny any dissociation between mindreading and metacognitive abilities, in line with what his theory predicts. It also avoids his having to make the hard-to-test claim that it is intermittently faulty data and not an intermittently faulty mechanism that is to blame for passivity symptoms.

Second, Carruthers holds that humans have introspective access to some mental states (e.g., perceptual states, imagery, and inner speech), but not to cognitive attitudes such as belief. But if information is extracted from globally broadcast perceptual states in third-person mindreading without introspection occurring, why think that the extraction of information from inner speech and visual imagery during first-person mindreading involves an introspective process different "in kind" from the way we form beliefs about the mental states of others? If, as Carruthers argues, passivity symptoms result from faulty data being input to the mindreading system (data that should have been interpreted as internally generated is interpreted as externally generated), then it seems the very determination of whether an input is self or other-generated – and thus whether one is seeing or visualizing, hearing or sub-vocalizing – requires an inferential or interpretative step (Langland-Hassan 2008).

Carruthers would likely respond that this inner-or-outer inferential step involves nothing more than the "trivial" form of inference that occurs in any layered representational scheme, where representations at one level can, in a "supervisory" role, intervene on those at another. However, many instances of third-person mindreading are equally fast and automatic, and they are implicit in the very cases of metacognition that, on Carruthers' theory, would be achieved through the "encapsulated" process of introspection. Consider a visual representation had by someone who looks up and sees another person staring at him. Suppose this visual perceptual state is accessed by the mindreading system, which issues in the introspective judgment: "I see a man seeing me." This judgment contains within it a judgment that another person is having a visual experience of a certain kind (cf. Jeannerod & Pacherie's [2004] "naked intentions"). So, unless the mindreading faculty in its introspective mode lacks the concepts needed for this judgment (unlikely, since it must have the concepts of self and of sight in order to issue *any* introspective judgments about visual experience), third-person mindreading can occur through the encapsulated "introspective" process that Carruthers describes. Yes, some cases of third-person mindreading require much more sophisticated feats of interpretation, but so too do many cases of first-person mindreading, as revealed by the confabulation data Carruthers discusses (Gazzaniga 1995).

Thus, even if it is possible to draw a line between mindreading that is informationally encapsulated and that which is not, it will not cut cleanly across cases of first- and third-person mindreading. Nor is the existence of such domain-specific mechanisms supported by recent neuroimaging studies (Decety & Lamm 2007). What we have instead are inferences, concerning both first- and third-person mental states, that require greater or lesser degrees of supporting information; none of this implies a special *mode* of access to facts about one's own mental states. This is obscured by the tendency of researchers to compare easy cases of metacognition (e.g., inferring one's intentions from one's own inner speech) with difficult cases of third-person mindreading (e.g., inferring what someone thinks based solely on their posture and facial expression) – for it creates the impression that first-person mindreading occurs through some more "direct" process. But if we instead compare the third-person mindreading that occurs when we judge that a person believes what we hear her saying, to the first-person mindreading that draws on "listening" to one's own inner speech, there is less intuitive pressure to posit a difference in the kind of inference. Of course, if there were genuine dissociations revealed between third- and first-person mindreading abilities, as Nichols and Stich (2003) and Goldman (2006) claim, then we would have reason to posit differences in the kinds of mechanisms and inferences involved in each; but Carruthers is at pains to deny any such dissociations, and his alternative explanations are plausible enough.

The issue can be reframed in terms of the larger evidence base we have for first-person rather than third-person mindreading. Carruthers notes that the resources available to first-person mindreading are different because, "unless subjects choose to tell me, I never have access to what they are imagining or feeling" (sect. 2, para. 8). This is potentially misleading; the situation is rather that the single mindreading system, as he describes it, *only ever* has access to globally broadcast perceptual and quasi-perceptual representations (and memory), and, with this single source of information, must accomplish both its first- and third-person mindreading tasks – one of which is to determine whether the signal counts as a case of imagining or perceiving in the first place.

The fact that we have so much more "evidence" for first-person mindreading than third-person may still tempt some to posit