How "weak" mindreaders inherited the earth

doi:10.1017/S0140525X09000570

Cameron Buckner,[a] Adam Shriver,[b] Stephen Crowley,[c] and Colin Allen[d]

[a]*Department of Philosophy, Indiana University, Bloomington, IN 47405-7005;* [b]*Philosophy-Neuroscience-Psychology Department, Washington University in St. Louis, St. Louis, MO 63130;* [c]*Department of Philosophy, Boise State University, Boise, ID 83725-1550;* [d]*Department of History and Philosophy of Science, Indiana University, Bloomington, IN 47405.*

cbuckner@indiana.edu
http://www.indiana.edu/~phil/GraduateBrochure/IndividualPages/cameronbuckner.htm ajshrive@artsci.wustl.edu
http://artsci.wustl.edu/~philos/people/index.php?position_id=3&person_id=60&status=1
stephencrowley@boisestate.edu
http://philosophy.boisestate.edu/Faculty/faculty.htm
colallen@indiana.edu
http://mypage.iu.edu/~colallen/

Abstract: Carruthers argues that an integrated faculty of metarepresentation evolved for mindreading and was later exapted for metacognition. A more consistent application of his approach would regard metarepresentation in mindreading with the same skeptical rigor, concluding that the "faculty" may have been entirely exapted. Given this result, the usefulness of Carruthers' line-drawing exercise is called into question.

Carruthers' recent work on metacognition in the target article (and in Carruthers 2008b) can be seen as an extended exercise in "debunking" metarepresentational interpretations of the results of experiments performed on nonhuman animals. The debunking approach operates by distinguishing "weak" metacognition, which depends only on first-order mechanisms, from "genuine" metacognition, which deploys metarepresentations. Shaun Gallagher (2001; 2004; with similar proposals explored by Hutto 2004; 2008) has been on a similar debunking mission with respect to metarepresentation in human mindreading abilities. Gallagher's position stands in an area of conceptual space unmapped by Carruthers' four models, which all presuppose that an integrated, metarepresentational faculty is the key to mindreading. Gallagher argues that most of our mindreading abilities can be reduced to a weakly integrated swarm of first-order mechanisms, including face recognition and an ability to quickly map a facial expression to the appropriate emotional response, a perceptual bias towards organic versus inorganic movement, an automated capacity for imitation and proprioceptive sense of others' movements (through the mirror neuron system), an ability to track the gaze of others, and a bias towards triadic gaze (I-you-target). Notably, autistic individuals have deficiencies throughout the swarm.

Someone pushing a "metarepresentation was wholly exapted" proposal might argue as follows: Interpretative propositional attitude ascription is a very recent development, likely an exaptation derived from linguistic abilities and general-purpose concept-learning resources. Primate ancestors in social competition almost never needed to think about others not within perceptual range; in the absence of language which could be used to raise questions and consider plans concerning spatially or temporally absent individuals, there would have been little opportunity to demonstrate third-person mindreading prowess. After developing languages with metarepresentational resources, our ancestors' endowment with the swarm would have left them well placed to acquire metarepresentational mindreading and metacognition through general learning. While such abilities were likely favored by cultural evolution in comparatively recent history, it is not clear that any further orders to genetic evolution needed to be placed or filled. Evolutionary "just so" stories come cheap; if Carruthers wants to make a strong case that the faculty evolved in response to social pressures (instead of just excellence with the swarm and/or other general aspects of cognition thought to be required for Machiavellian Intelligence, such as attention, executive control, and working memory), he needs further argument.

Two issues must be overcome for the swarm proposal to be considered a serious alternative. First, the concurrent appearance of success on verbal first- and third-person false-belief tasks must be explained. Here, we point the reader to Chapter 9 of Stenning and Van Lambalgen (2008), which makes a strong case that the logic of both tasks requires a kind of conditional reasoning which does not develop until around age 4 and is also affected by autism (and see also Perner et al. [2007] for a related account). Second, there is the work on implicit false-belief tasks with prelinguistic infants (Onishi & Baillargeon 2005). These findings are both intriguing and perplexing (consider, for example, that the infants' "implicit mastery" at 15 months is undetectable at 2.5 years), and the empirical jury is still out as to whether the evidence of preferential looking towards the correct location can support the weight of the metarepresentational conclusions which have been placed on it (see Perner & Ruffman 2005; Ruffman & Perner 2005). The infants' preferential looking can be explained if they quickly learn an actor-object-location binding and register novelty when the agent looks elsewhere. More recent studies (e.g., Surian et al. 2007) claiming to rule out alternatives to the metarepresentational explanation have produced findings that are ambiguous at best (Perner et al. 2007).

One might concede that the mechanism generating the gaze bias in infants is not itself metarepresentational, but nevertheless hold that it evolved because it enabled its possessors to develop metarepresentation – likely wielding a poverty of the stimulus (PoS) argument to the effect that even with language, metarepresentational mindreading does not come for free. We suggest that such reasoning no longer carries the weight it once did. Recent work on neural network modeling of the hippocampus, which highlights its ability to quickly discover abstract, informationally efficient bindings of stimulus patterns (especially when fed neutral cues like words – e.g., see Gluck & Myers 2001; Gluck et al. 2008) dulls the PoS sword. Finally, even if the PoS argument is accepted, there remains a huge leap to the conclusion that the bias evolved *because of its ability to bootstrap metarepresentation* – and not for something simpler.

In light of the swarm alternative, the usefulness of Carruthers' distinction between "weak" and "genuine" forms of mindreading and metacognition becomes questionable. Our overarching worry is that Carruthers' emphasis on a single faculty of metarepresentation, combined with his acknowledgment of the rich heritage of cognitive abilities shared between humans and animals, leaves the faculty almost epiphenomenal in human cognition (except, perhaps, for Machiavelli himself) – a position that Carruthers has previously been driven to adopt with respect to his account of phenomenal consciousness (Carruthers 2005; see also Shriver & Allen 2005). An alternative approach might be to tone down the deflationary invocation of first-order mechanisms, and focus instead on what creatures endowed with a swarm of weakly integrated mechanisms can do and learn. Once we abandon the assumption that mindreading is centralized in a single metarepresentational faculty, we can investigate whether something like Gallagher's swarm could implement various degrees of competence in reacting adaptively to the mental states of others. This perspective focuses us on the flexibility and adaptive significance of the evolved mechanisms which

constitute the swarms, for a wide range of organisms in a variety of social environments (including humans in theirs). These suggestions are in the spirit of Dennett (1983), who advocated the usefulness of metarepresentational hypotheses in devising new experiments, accepting from the beginning that animals and humans will "pass some higher-order tests and fail others" (p. 349). Ultimately, we think that the questions Carruthers raises about the relationship between self-regarding and other-regarding capacities are interesting and should be pursued; and they *can* be pursued without engaging in the line-drawing exercise which de-emphasizes the significance of good comparative work for understanding human cognition.

ACKNOWLEDGMENT

We thank Jonathan Weinberg for his extensive comments on earlier versions of this commentary.