4-1-2009

How "Weak" Mindreaders Inherited the Earth

Cameron Buckner
Indiana University - Bloomington

Adam Shriver
Washington University in St. Louis

Stephen Crowley
Boise State University

Colin Allen
Indiana University - Bloomington

That is true for humans, but it is equally true for animals, who must survive real-world challenges in environments in which errors lead to extinction. Brain evolution is not separate from the ability to observe and know the real world. On the contrary, when we are given truthful feedback about the world, humans and other animals become quite reality-based. There is no contradiction between constructivism and realism.

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.

Cognitive science at fifty

doi:10.1017/S0140525X09000582

A. Charles Catania
Department of Psychology, University of Maryland, Baltimore County (UMBC), Baltimore, MD 21250.
catania@umbc.edu
http://www.umbc.edu/psyc/personal/catania/catanias.html

Abstract: Fifty years or so after the cognitive revolution, some cognitive accounts seem to be converging on treatments of how we come to know about ourselves and others that have much in common with behavior analytic accounts. Among the factors that keep the accounts separate is that behavioral accounts take a much broader view of what counts as behavior.

Roughly half a century has passed since the cognitive revolution declared behaviorism dead and promised solutions to long-standing problems of philosophy and psychology. Carruthers provides an opportunity to assess the progress that has taken place. Mind remains central in his account, and its hierarchical structure is illustrated in the pivotal roles of metarepresentations and metacognitions. In place of behavior and events in the world, the action takes place in the dynamics of their surrogates, such as perceptions and intentions and beliefs and concepts and attitudes, none of which lend themselves to measurement in the units of the physical or biological sciences. Most of the entities in Carruthers' account existed in the vocabularies of the mid-1950s, though typically more closely anchored to their origins in colloquial talk, which since then has sometimes been called folk psychology.

What has most obviously changed are the linkages among the mentalistic terms. Carruthers deals with the particular priorities of mindreading and metacognition. Are they independent mechanisms or a single mechanism with two modes of access? Is one a prerequisite for the other? Carruthers concludes that metacognition is grounded in mindreading. If one argues that judgments about oneself must be distinguished from judgments about others, his conclusion is sound. But this conclusion is one that a variety of behaviorism reached long before the advent of the cognitive revolution. In his "Behaviorism at 50," Skinner (1963) recounted the history of Watsonian methodological behaviorism in the early decades of the twentieth century and its rejection of introspection (see also Catania 1993), but he also noted the unnecessary constraints that Watson's account had imposed on theory.

Skinner's later radical behaviorism rejected the Watsonian constraints and extended his approach to the origins of the language of private events. As a contribution to a symposium organized by his advisor, E. G. Boring, Skinner (1945) made explicit his interest in "Boring from Within." The 1945 paper, "The Operational Analysis of Psychological Terms," was a renunciation of operationism, but, more important, it provided an account of how a vocabulary of private events (feelings, emotions, etc.) could be created even though those who taught the words and maintained consistencies of usage had access only to shared public accompaniments of those private events.[1] Given these origins of the private or introspective language, Skinner's resolution of the issue in terms of the public practices of the verbal community is the only feasible way of dealing with the problem that Carruthers has so aptly described in terms of his mindreading system, which never has access to what others are imagining or feeling. To the extent that it does have access to what one feels or imagines oneself, one can speak of those events only in a vocabulary that is anchored in public correlates. Carruthers' point that instances of self-attributed unsymbolized thought occur in circumstances in which a third party might have made the same attribution is perfectly consistent with this argument.

The irony, then, is that with respect to introspection, judgments about the behavior of others (mindreading) and judgments about one's own behavior (metacognition), Carruthers has reached conclusions that are consistent with Skinner's. One can guess that he took so long only because of the complexity of the terms that entered into his account. Skinner's account is far more parsimonious. Skinner does not begin with something called discriminating and follow it with differential responding; the differential responding is itself the discriminating. He does not say that perceiving and sensing and thinking are something different from behaving; they are kinds of behavior, defined not by whether they involve movement but rather by whether they are involved in contingent relations with environmental events (for this reason, Carruthers notwithstanding, a lot of behavior goes on even when one is sitting quiet and motionless, and one has just as much access to this behavior as to that of standing or walking). There is no more need to appeal to seeing and hearing as prerequisite concepts than there is to say that we cannot sit or stand or walk without concepts of sitting or standing or walking; these are all names for things we do. To invoke them as explanations does not serve our theories well.

Carruthers' account also converges on other concepts that have been elaborated by Skinner. For example, his System 1 and System 2 have features that are closely paralleled by what Skinner (1969) respectively called rule-governed and contingency-shaped behavior, and Carruthers is surely on the right track in saying that speech is an action that does not begin with metacognitive representations of thought (a more detailed account is beyond the scope of this commentary, but see Catania 2006, Chs. 14 and 15). Furthermore, in considering the different environmental contingencies that operate on verbal and nonverbal classes of behavior, the behavioral account has no trouble dealing with the various confabulations that Carruthers has surveyed. Just as speech errors can tell us a lot about language structure, so confabulations may tell us a lot about the nature of our judgments about ourselves and others.

It is good to see cognitive science at last converging on conclusions that had once been reached in behavioral accounts. If that were the only point, this commentary would serve little but a historical purpose. But there is extensive behavior analytic research relevant to these issues (in particular, see Wixted & Gaitan 2002), and some of it may prove useful to those of any theoretical orientation. Of course, it would be not at all surprising if the suggestions here are not well received. That likelihood is enhanced by the fact that this has been a necessarily brief and superficial presentation of the behavioral case. But the literature is there, so perhaps a few will check it out.

NOTE

[1] Two articles by B. F. Skinner cited in this commentary (Skinner 1945; 1969) were reprinted in *Behavioral and Brain Sciences* (Vol. 7, December 1984).