Review

The study of animal metacognition

J. David Smith

Department of Psychology and Center for Cognitive Science, University at Buffalo, the State University of New York, 346 Park Hall, SUNY Buffalo, Buffalo, NY 14260, USA

Do nonhuman animals share humans' capacity for metacognition—that is, for monitoring or regulating their own cognitive states? Comparative psychologists have approached this question by testing a dolphin, pigeons, rats, monkeys and apes using perception, memory and food-concealment paradigms. There is growing evidence that animals share functional parallels with humans' conscious metacognition, although the field has not confirmed full experiential parallels and this remains an open question. This article reviews this new area of comparative inquiry and describes significant empirical milestones, remaining theoretical millstones and the prospects for continuing progress in a rapidly developing area. This research area opens a new window on reflective mind in animals, illuminating its phylogenetic emergence and allowing researchers to trace the antecedents of human consciousness.

Introduction

Humans can feel uncertainty. They know when they do not know or remember, and they respond well to uncertainty by deferring response and seeking information. These adaptive responses ground wide-ranging research on metacognition, which means thinking about thinking [1–6]. The organizing idea in this field is that human minds have a cognitive executive or executive function that monitors and controls perception and memory.

Metacognition is a sophisticated human capacity [7]. It is linked to hierarchical structure in the mind because the higher-level executive function oversees lower-level cognition [8], to self-awareness because uncertainty and doubt are so personal and subjective [9], and to declarative consciousness because humans are conscious of their states of knowing and can declare them to others [4,10]. This sophistication might imply that metacognition is uniquely human. Therefore, one of comparative psychology's current goals is to establish whether nonhuman animals (hereafter, animals) share humans' metacognitive capacity [11]. If they do, it could bear on their consciousness and self-awareness too. Metacognition rivals language and tool use in its potential to establish important continuities or discontinuities between human and animal minds.

This article reviews this rapidly growing area of comparative inquiry. First, I describe the early perceptual studies that demonstrated animals' uncertainty responses. The metacognitive interpretation of these studies is debated given alternative interpretations. Second, I describe recent studies that tested the alternative interpretations. These studies strengthen the evidence that animals and humans have functionally parallel metacognitive systems. Third, I consider the appropriate psychological interpretation of the performances animals have achieved. Fourth, I consider whether animals' metacognitive capacity is a declarative cognitive process imbued with conscious awareness. Taken as a whole, the comparative-metacognition field offers growing evidence that some animals have functional analogs to human consciousness, and suggests that researchers may be opening an empirical window on animals' cognitive awareness.

Perceptual studies of uncertainty responding

Traditional measures of human metacognition (e.g. verbalized feelings of knowing, self-reported tip-of-the-tongue states) are not suitable for exploring animal metacognition. These measures are too verbal, too introspective and too dependent on self-report for subjects who cannot respond 'Wait, wait, don't tell me!'. Accordingly, early studies in this area adopted perceptual–behavioral tasks with two components. First, animals were given difficult perceptual discriminations: the difficulty potentially created uncertainty in their minds. Second, animals were given an additional response—beyond the discrimination responses—with which they could decline to complete any trials of their choosing. This response—sometimes called the Uncertainty Response (UR)—allows animals to report on, or cope with, the difficulty. If animals monitor cognition accurately, they should prospectively recognize difficult trials as error-risking and decline those trials selectively.

For example, macaques in [12] performed a Sparse-Dense discrimination. They used a joystick to move a cursor to one of three objects on a computer screen (Figure 1). Touching the cursor to the Box or S, respectively, was correct if the box contained 2,950 pixels (Dense) or any fewer pixels (Sparse). They received food and time-outs, respectively, for correct and incorrect responses. The UR (touching the Star) let subjects decline the present trial and enter a guaranteed-win trial in which a Box or an S was presented alone. The subjects were rewarded if they touched the presented object. Monkeys responded correctly to easy Sparse trials and to many Dense trials (Figure 2a). They were at chance where these response curves crossed. Most important, they assessed correctly when they risked a discrimination error and they made URs to selectively decline these trials.

Humans performed similarly (Figure 2b). There is a strong cross-species isomorphism in the use of the UR (Box 1, [13]) that produces some of the closest existing human–animal performance correspondences. Moreover, humans report that their URs reflect conscious uncertainty. Given the self-reports and the performance similarities,

Corresponding author: Smith, J.D. (psysmith@buffalo.edu).

1364-6613/$ – see front matter doi:10.1016/j.tics.2009.06.009 Available online 31 August 2009

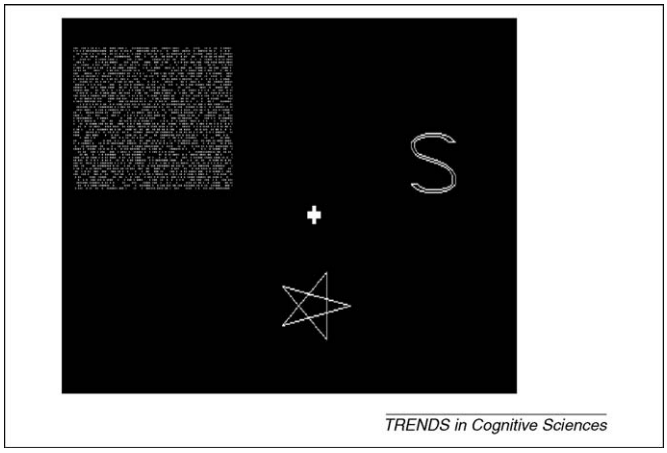

Figure 1. The screen from a trial in a Sparse-Dense discrimination [12]. From 'The uncertain response in humans and animals,' by J.D. Smith, W.E. Shields, J. Schull, and D.A. Washburn, 1997, *Cognition, 62,* p. 79. Copyright 1997 by Elsevier. Reprinted with permission.

should one extend the psychological interpretation of conscious uncertainty to monkeys too? If monkeys' URs do demonstrate animals' capacity for conscious uncertainty monitoring and metacognition, then they are important behavioral ambassadors that reveal reflective mind and awareness. However, the next section explains why one must extend the psychological interpretation cautiously.

Theoretical concerns

A tradition of parsimony leads comparative psychologists to interpret an organism's behavior at the lowest possible psychological level [14]. Thus, even given animal performances that seem metacognitive, they suppose that the performances might be explained using low-level, associative/conditioning mechanisms. Some animal metacognition studies do raise this alternative possibility.

First, consider that animals have often been given concrete rewards for URs [13,15–20]. This methodology could grant the UR a general attractiveness independent of its uncertainty-monitoring role. Its use might reflect its reward properties, not a metacognitive judgment. This methodology makes it difficult to rule out low-level interpretations or affirm metacognitive interpretations.

Second, perceptual studies focused on stimulus qualities (box density, tone pitch) and thus encouraged stimulus-based performance descriptions. It might be that concrete stimuli, not introspected metacognitive states, trigger URs. In particular, subjects could develop aversion to the stimuli that cause errors/timeouts and avoid responses to those stimuli, whereas URs to those stimuli were safe and attractive.

Third, the trial-by-trial reinforcement usually given encourages reinforcement-based performance descriptions. Animals might track their reinforcement history for different trial types and maximize reinforcement through URs on the least-reinforced types. Metacognition might not enter into it. The formal models in [21] concretized this possibility.

Fourth, animals are usually extensively trained in uncertainty-monitoring tasks. They might become sensitive to some task-specific cue to stimulus difficulty, perhaps even one unknown to the experimenters. URs could be occasioned by this cue and not by a generalized state of uncertainty.

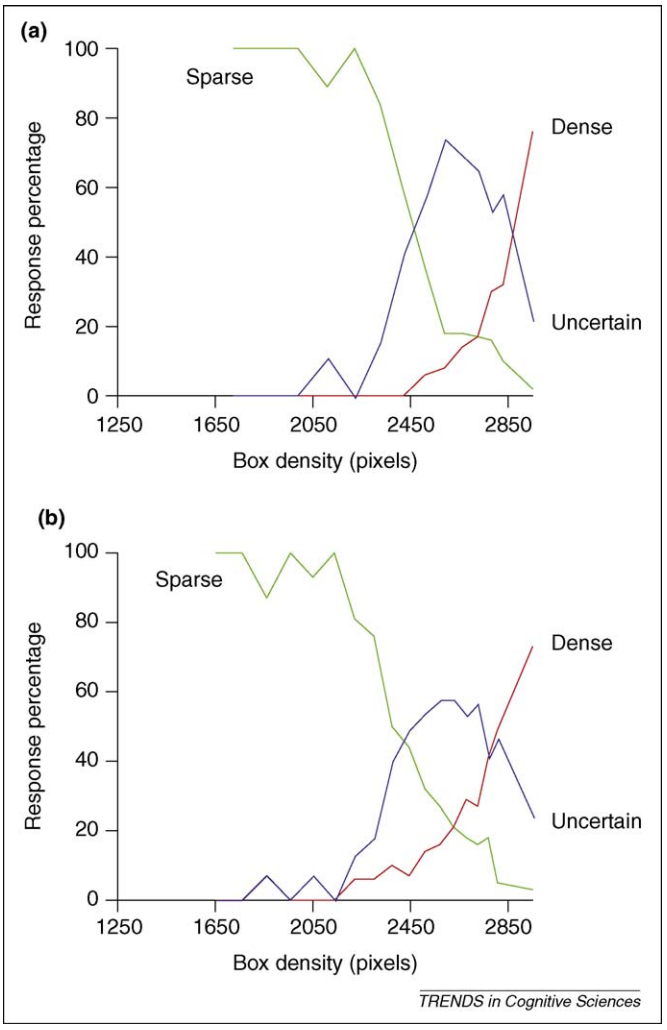

Figure 2. **(a)**. Performance by a monkey in a Sparse-Dense discrimination [12]. The horizontal axis indicates the density of the trial. The Dense response was correct for 2,950-pixel trials—these trials are represented by the rightmost data point for each curve. All trials with fewer pixels deserved the Sparse response. The blue line represents the percentage of trials receiving the uncertainty response at each trial level. The percentages of trials ending with the Sparse response (green line) or Dense response (red line) are also shown. **(b)** The performance of humans in the Sparse-Dense discrimination, depicted in the same way. To equate discrimination performance across subjects, the data were normalized to place each subject's discrimination crossover at a pixel density of about 2700. From 'The Comparative Psychology of Uncertainty Monitoring and Metacognition,' by J.D. Smith, W.E. Shields, and D.A. Washburn, 2003, *Behavioral and Brain Sciences, 26,* p. 322. Copyright 2003 by the Cambridge University Press. Reprinted with permission.

These four aspects of the original studies left open the question of whether animals' performances showed metacognition or some low-level conditioning mechanism.

Stage 2 animal metacognition studies

Facing this question, researchers developed new paradigms to distance animals' performance from the reward properties of URs, from the stimulus cues that could trigger avoidance responses, from the reinforcement contingencies that could motivate avoidance and from the overtraining that could grant animals access to task-specific difficulty cues.

Pure URs

In [22,23], macaques still made adaptive URs when those responses brought no hint or information, no easy next trial, and no food reward or reward token, but only the