BEHAVIORAL AND BRAIN SCIENCES (2003) **26,** 317–373
Printed in the United States of America

The comparative psychology of uncertainty monitoring and metacognition

J. David Smith
Department of Psychology and Center for Cognitive Science, State University of New York at Buffalo, Buffalo, NY 14260.
psysmith@buffalo.edu
http://wings.buffalo.edu/psychology/labs/smithlab/

Wendy E. Shields
Department of Psychology, University of Montana, Missoula, MT 59812.
wendy.shields@umontana.edu
http://psychweb.psy.umt.edu/faculty/shields/shields.html

David A. Washburn
Department of Psychology and Language Research Center, Georgia State University, Atlanta, GA 30303.
dwashburn@gsu.edu
http://www.gsu.edu/~wwwpsy/faculty/washburn.htm

Abstract: Researchers have begun to explore animals' capacities for uncertainty monitoring and metacognition. This exploration could extend the study of animal self-awareness and establish the relationship of self-awareness to other-awareness. It could sharpen descriptions of metacognition in the human literature and suggest the earliest roots of metacognition in human development. We summarize research on uncertainty monitoring by humans, monkeys, and a dolphin within perceptual and metamemory tasks. We extend phylogenetically the search for metacognitive capacities by considering studies that have tested less cognitively sophisticated species. By using the same uncertainty-monitoring paradigms across species, it should be possible to map the phylogenetic distribution of metacognition and illuminate the emergence of mind. We provide a unifying formal description of animals' performances and examine the optimality of their decisional strategies. Finally, we interpret animals' and humans' nearly identical performances psychologically. Low-level, stimulus-based accounts cannot explain the phenomena. The results suggest granting animals a higher-level decision-making process that involves criterion setting using controlled cognitive processes. This conclusion raises the difficult question of animal consciousness. The results show that animals have functional features of or parallels to human conscious cognition. Remaining questions are whether animals also have the phenomenal features that are the feeling/knowing states of human conscious cognition, and whether the present paradigms can be extended to demonstrate that they do. Thus, the comparative study of metacognition potentially grounds the systematic study of animal consciousness.

Keywords: cognition; comparative cognition; consciousness; memory monitoring; metacognition; metamemory; self-awareness; uncertainty; uncertainty monitoring

1. Introduction

Extensive research on metacognition and uncertainty monitoring has been done to explore humans' capacity to recognize uncertainty and to know when they do not know (Brown et al. 1982; Dunlosky & Nelson 1997; Flavell 1979; Koriat 1993; Metcalfe & Shimamura 1994; Nelson 1992; Nelson & Dunlosky 1991; Nelson & Narens 1990; Reder 1996; Schwartz 1994). Human adults and older human children (hereafter, humans) respond adaptively when facing difficult or uncertain situations – they defer response and seek help, hints, or additional information.

Less research has been done to explore the metacognitive capacities of nonhuman animals (hereafter, animals – Hampton 2001; Inman & Shettleworth 1999; Shields et al. 1997; Smith et al. 1995; 1997; 1998; Smith & Schull 1989; Teller 1989). We consider the present status and future prospects of this area of comparative psychology. We believe the area would benefit if interested colleagues evaluated existing research and guided future research through their commentaries.

The article proceeds as follows. First, we describe a theoretical framework for metacognition. Second, we discuss potential contributions from a comparative psychology of metacognition. Third, we discuss the requirements for comparative metacognition paradigms that have caused this field to develop slowly. Fourth, we consider existing uncertainty-monitoring results in the domains of perception (Shields et al. 1997; Smith et al. 1995; 1997) and memory (Hampton 2001; Smith et al. 1998). Fifth, we extend phy-

© 2003 Cambridge University Press 0140-525X/03 $12.50

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

logenetically the search for metacognitive capacities in animals by considering studies that tested less cognitively sophisticated species (rats and pigeons – Inman & Shettleworth 1999; Smith & Schull 1989; Teller 1989). It is a potentially important fact that these species have not shown evidence of metacognitive capacities, whereas humans, monkeys, and a dolphin have. Sixth, we provide a unifying formal description of performance in the tasks herein described. Seventh, we discuss the appropriate psychological interpretation of animals' performances. Eighth, we consider the nettlesome relationship between these uncertainty-monitoring performances and the declarative consciousness of uncertain mental states.

2. A theoretical framework for metacognition

Metacognition is defined to be cognition about cognition. The idea in this field is that some minds contain a cognitive executive that looks in on thought or problem solving to see how it is going and how it might be facilitated (e.g., as when we realize that a paragraph of an article has not been understood and reread it). Nelson and Narens (1990) gave the literature on human metacognition a useful theoretical framework (Fig. 1). They theorized that mental activities occur at a meta level and at a lower, object level during cognitive processing. The meta level monitors the processing and determines its progress and prospects. These monitoring functions of the meta level (i.e., the basic metacognitive judgments) are shown at the top of Figure 1. There is an ease-of-learning judgment about whether material will be easy or hard to learn (pepperoni-pizza vs. pepperoni-thesaurus). There is a judgment of learning about the level of learning achieved. There is a feeling-of-knowing judgment about whether information is potentially available in memory or not (e.g., the middle names of William___Clinton vs. Anthony___Blair). There is a judgment about one's confidence in the accuracy of retrieved answers.

The meta level also controls cognitive processing, directing information processing in ways that may be more felicitous. These metacognitive control processes are shown at the bottom of Figure 1. The meta level may select new processes (e.g., elaborative rehearsal instead of maintenance), allocate extra study time to difficult items, terminate studying when sufficient learning is judged to have been achieved (some undergraduates make this judgment too optimistically), select new retrieval strategies when present ones are failing, or abandon retrieval efforts if success seems improbable.

Figure 1 lays out an ambitious potential research program that could show all of these capacities in animals but which has barely begun. Thus far researchers have evaluated whether animals monitor the certain or uncertain status of ongoing perception and memory. No studies have considered whether animals use their uncertainty to alter adaptively the character of information processing. So there remain interesting lines of research to be pursued regarding both the metacognitive control processes and the metacognitive monitoring processes shown in Figure 1.

3. Potential contributions of a comparative psychology of metacognition

The comparative study of metacognition potentially illuminates important issues in comparative psychology and cognitive science. First, metacognition is considered one of humans' most sophisticated cognitive capacities and possibly a uniquely human cognitive capacity. It is an important question whether this capacity extends to other species. Second, the comparative study of metacognition would expand the study of animal self-awareness that has depended on the elegant but controversial mirror dye-mark test that assays animals' bodily self-awareness (Gallup 1982; Gallup & Suarez 1986; Parker et al. 1994; Swartz 1997; Swartz et al. 1999). Direct measures of cognitive self-awareness, which may be a different thing from self-recognition (Cheney & Seyfarth 1990, p. 240), would be a useful addition to this area.

Third, comparative metacognition research would contribute to theory-of-mind research. Theory-of-mind research asks whether animals know and monitor the other's mental states and states of knowing (Byrne & Whiten 1991; Cheney & Seyfarth 1990; Heyes 1998 and associated commentaries; Whiten & Byrne 1997). The complementary question is whether animals know and monitor their own mental states and states of knowing (Schull & Smith 1992). The relation between these capacities is an important issue in discussions about theory of mind and the evolution of social intelligence. Some theories stress the evolutionary interdependence between self- and other- mental awareness. Perhaps self-awareness evolved in social species to facilitate other-awareness and social intelligence (see Humphrey 1976). Indeed, perhaps self-awareness was a prerequisite for other-awareness. Direct measures of cognitive self-awareness could let theorists explore such possibilities.

John David Smith is Associate Professor in the Departments of Psychology and Music at the University at Buffalo, State University of New York. He received his Ph.D. from the University of Pennsylvania in 1982. He studies uncertainty monitoring and metacognition, the psychology of categorization, the perceptual problems faced by security screeners at airports, and the psychology of music and aesthetics. His research is supported by the National Institute of Child Health and Human Development and the Transportation Security Administration.

Wendy E. Shields is Assistant Professor in the Animal Behavior and Cognition program at the University of Montana, a position she has held since 1999. She received her Ph.D. in cognitive psychology from the University at Buffalo, State University of New York. Her goal in her current research is to gain a better understanding of metacognitive or metacognitive-like processes by studying them comparatively across species.

David A. Washburn is Associate Professor in the Department of Psychology and Director of the Language Research Center at Georgia State University (Atlanta, Georgia, USA). He received his Ph.D. from that institution in 1991. His research is focused on individual and group (including species) differences in attention and executive functioning, including planning, monitoring, set-switching, and inhibition. His research enjoys support from numerous agencies including the National Institute of Child Health and Human Development, the Federal Aviation Administration, the U.S. Army, and the McDonnell-Pew Foundation.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

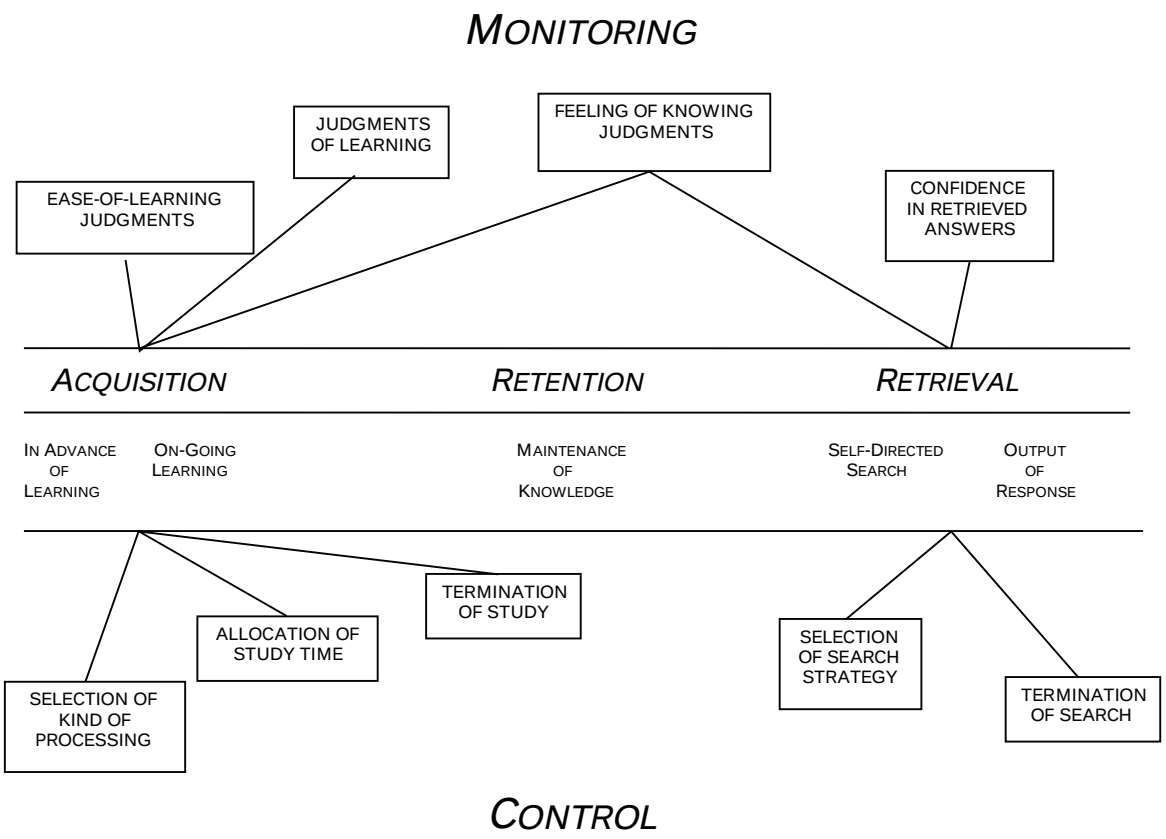


Figure 1. A theoretical framework for research on metacognition, showing examples of process-monitoring capacities above and process-control capacities below (after Nelson & Narens 1990).

Fourth, given the link between humans' metacognition and declarative consciousness (Nelson 1996), the study of animal metacognition would contribute to the study of animal consciousness. Weiskrantz (1986; 1997, Ch. 4, pp. 77–99) discussed the possibility of studying animal consciousness through behavioral responses. In his thought experiment, animals had available two discrimination responses and also a commentary key with which to step outside the discrimination and report on the state of their knowledge or perception. The tasks described in this article provide to animals this commentary key in a variety of settings. We agree that these tasks suggest the possibility of exploring animal consciousness systematically. However, we will discuss the difficulty of inferring declarative consciousness from animals' performances in tasks of this kind. We will also discuss the possible theoretical separation between animals' having the functional parallels of humans' conscious cognition and animals' having the subjective, phenomenal feelings of doubt and knowing.

Fifth, because the same uncertainty-monitoring paradigms can be used across several species, metacognitive capacities can be assayed comparably in different phylogenetic groups. This means that it should eventually be possible to draw the map of the phylogenetic distribution of metacognition or cognitive self-awareness. This map might illuminate the emergence of mind.

Sixth, species differences in these capacities might reveal the cognitive mechanisms underlying metacognition. For example, comparing the capacity for uncertainty monitoring in species with and without language could indicate the role of language in metacognition. Correlating uncertainty awareness with the capacities for planning and future-oriented cognition could show the extent to which metacognition is prospective and allied to other executive functions. The comparative psychology of metacognition could also suggest the earliest roots of metacognition in human development and provide the techniques for investigating them. In these and other ways we believe the parallel investigation of animal and human metacognition could produce a constructive synergy.

4. Requirements for a comparative uncertainty paradigm

There are two basic requirements for creating a comparative uncertainty paradigm. One is to create perceptual or cognitive difficulty for the animal in order to stir up something like an uncertainty state. The other is to provide a behavioral (i.e., nonverbal) response that lets the animal comment on or cope adaptively with that state. (This second requirement explains why comparative metacognition research began slowly. The typical human paradigms did not suit animals, for their phenomena relied heavily on verbal self-reports about feelings of knowing, judgments of learning, tip-of-the-tongue experiences, and so forth – Brown 1991; Hart 1965; Smith et al. 1991.)

These two requirements can be illustrated by considering psychophysical procedures. These procedures are highly developed for creating difficulty for animals. They

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

narrow the contrast between alternative stimulus input classes and force observers to make difficult discriminations at their perceptual thresholds (Au & Moore 1990; Blough 1958; Schusterman & Barrett 1975; Yunker & Herman 1974). They always cause the animal difficulty. The important question for metacognition research is whether animals sense this difficulty and could respond adaptively to it – that is, whether psychophysical procedures generate useable uncertainty states in animals.

Causing the animal difficulty is not sufficient for answering this question. Illustrating this point, Figure 2 shows the performance of a dolphin (*Tursiops truncatus*) near threshold in a discrimination between 2,100-Hz tones and tones at any lower frequency (Smith et al. 1995). The animal made one response (Low) correctly to trials below about 2,075 Hz and the other response (High) correctly to many 2,100-Hz trials. The trials surrounding 2,085 Hz, his known threshold relative to 2,100 Hz (Herman & Arbeit 1972), produced near-chance performance. This task causes the animal the difficulty it should. It might be creating the uncertainty states the comparative metacognition researcher seeks to study. But it cannot show whether the animal senses the difficulty or could cope with the uncertainty. These capacities are hidden by allowing only two responses that map to the two input classes (2,100-Hz tones and lower tones) and by denying the animal any way to comment on uncertainty or respond adaptively to it.

It is possible that animals' ancillary behaviors would betray their uncertainty or conflict on threshold trials. These hesitations and waverings do sometimes occur, and they made behaviorists uncomfortable because they suggested that animals might be in mental turmoil over difficult trials. Tolman even suggested that these uncertainty behaviors – "lookings or runnings back and forth" – could be taken as a behaviorist's definition of animal consciousness (Tolman 1927). As a behaviorist, Tolman sometimes misbehaved. Moreover, 12 years later, Tolman (1938, p. 27) retracted this claim, thereby completing his own theoretical "looking or running back and forth." (See also Tolman 1932/1967.)

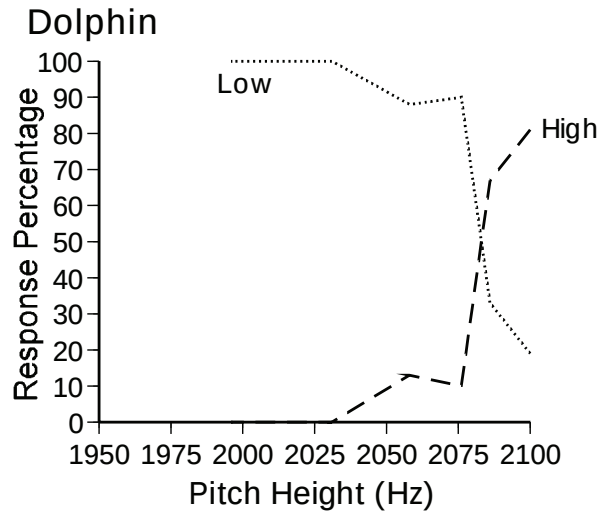


Figure 2. Frequency discrimination performance by a bottlenosed dolphin (*Tursiops truncatus*) in the procedure of Smith et al. (1995). The horizontal axis indicates the frequency (Hz) of the tone. The high response was correct for tones at exactly 2,100 Hz, and these trials are represented by the rightmost data point for each curve. All other tones deserved the low response. The percentages of trials ending with the high response (dashed line) or low response (dotted line) are shown.

However, it may not be advisable to rely on ancillary behaviors to convey information about the animal's uncertainty processes. These behaviors may not happen, they may be poorly interpretable or measurable, and they may defeat comparative research because animals in different species may react differently when facing uncertainty. For example, unlike uncertain humans, uncertain dolphins do not scratch their heads. Accordingly, the safest course in comparative metacognition research is to give animals of different species the same concrete response that lets them report on or deal with the difficult situation. This is the second requirement of a comparative uncertainty paradigm, to provide animals with a third, uncertain response that lets them cope with difficulty by declining the trials they do not choose to complete. Then a reasonable strategy for the animal is to use the uncertain response sparingly, when errors on the primary perceptual or cognitive task are judged to be likely. To carry out this strategy, animals must identify, if they can, the occasions on which they are liable to err.

5. The "Uncertain" response in human psychophysics

One methodology for studying animals' uncertainty monitoring combines psychophysical procedures that cause animals perceptual difficulty with the uncertainty response that lets them report on and cope with their uncertainty (Smith et al. 1995; 1997). This paradigm has a rich history in human experimental psychology that raises the possibility of studying metacognition comparatively.

Human observers in early psychophysical studies were often allowed to respond "Uncertain" when they felt unable to assign a stimulus to one of the two primary input classes (Angell 1907; Fernberger 1914; 1930; George 1917; Watson et al. 1973; Woodworth 1938). Some researchers hoped that the level of uncertainty responses would index sensory sensitivity (e.g., Urban 1910), with dull and sharp perceivers needing that response more and less, respectively. However, others questioned this approach (Fernberger 1914; Woodworth 1938, pp. 419–27), noting the special psychological status of the response "Uncertain" (henceforth Uncertain response). That response had longer latencies (Angell 1907; George 1917; Woodworth 1938) and special susceptibility to instructions (Brown 1910; Fernberger 1914; 1930; Woodworth 1938). It seemed to be linked to participants' personality or temperament (Angell 1907; Fernberger 1930; Thomson 1920) – a linkage that surely had little to do with studying sensory capability. It seemed to be more reflective and cognitive than the two primary discrimination responses (Angell 1907; Fernberger 1930; George 1917).

Theorists also had a structural concern about uncertainty responses. They believed that these were meta to the primary discrimination and were a comment on the participant's failure to assign a stimulus to one of the primary input classes. George (1917) concluded that the doubtful responses were offenders against the constant attitude required in psychophysics because they introduced "extra-serial" attitudes into a task that depended on intra-serial, sen-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

sory attitudes. Boring (1920) suggested that doubt was an attitudinal seducer that took observers away from the series of mental states that are a continuous function of the series of stimuli. Jastrow (1888) and Brown (1910) recommended forcing a primary discrimination response from observers on each trial, while collecting confidence judgments on the side to replace the information provided by uncertainty responses. Confidence ratings are often collected from humans in this way. The catch is that animals have so far not proved able to report their confidence in this way (but see Shields et al., submitted).

However, we will see that animals can respond "Uncertain." Given that fact, the structural problem early theorists had with this response makes it intriguing regarding a comparative psychology of uncertainty monitoring. The Uncertain response might be meta to animals' primary discriminatory process, too. It might be for them, too, a comment on or a response to indeterminacy and difficulty in the primary discrimination. There is contemporary convergence on this possibility. The psychophysical Uncertain response instantiates the commentary key that Weiskrantz (1986; 1997) considered providing to blindsight animals in a thought experiment. Cowey and Stoerig (1992) proposed a similar *Gedanken* procedure (see also Cowey & Stoerig 1995). Their idea was to train a light/no-light discrimination, then use psychophysical procedures to bring the animal to a threshold where it would be only 50% correct. Critically, though, the animal would also have a third lever reinforced on a 75% schedule. Then the adaptive animal could report on its state of not knowing whether a light was seen, by choosing the third lever selectively on threshold trials.

These converging ideas made it clear that the psychophysical uncertainty paradigm was a strong starting place for the comparative study of uncertainty monitoring. We discuss now the results when humans, monkeys, and a dolphin were placed into difficult and uncertain perceptual situations and forced to make judgments at threshold, but were also allowed to respond Uncertain.

As we discuss these results, readers will naturally consider both metacognitive interpretations of performance and possible lower-level interpretations of performance. For example, there might be inadvertent cueing of the animal (e.g., by a dolphin trainer) to respond Uncertain on difficult trials. Or, the animal might respond Uncertain to avoid aversive, error-producing stimuli, instead of doing so to cope adaptively with uncertainty. Additional research in this field (see sects. 8–10) has addressed some alternative accounts. Alternative accounts are also discussed in section 14 on the problem of psychological interpretation. Section 14.2 specifically discusses the role that low-level, associative explanations of behavior have in explaining comparative data.

6. Uncertain responses by humans and monkeys in a psychophysical density discrimination

To begin the comparative study of uncertainty monitoring, Smith et al. (1997) placed humans and rhesus monkeys (*Macaca mulatta*) in a visual density-discrimination task. Participants used a joystick to move a cursor to one of three objects on a computer screen (Fig. 3A). The dense response (choosing the box) was correct if the box contained exactly 2,950 illuminated pixels. The sparse response (choosing the S) was correct if the box had any fewer pixels. The Uncertain response (choosing the star) allowed participants to decline the trial and move into a new, guaranteed-win one. Initially, participants were stabilized on an easy discrimination involving 2,950- and 450-pixel boxes. Then, the discrimination's difficulty was gradually increased by making the sparse boxes denser until performance faltered at about 2,950 versus 2,600 pixels. At mature performance, trial difficulty was continuously adjusted based on participants' performance within a session to maintain difficulty at a constant, high level.

Figure 3B shows the performance of seven humans. Sparse responses predominated on sparser trials; dense responses predominated on true dense trials and the most difficult sparse trials. The primary discrimination was performed at chance where these two response curves cross. Humans responded Uncertain most in the region of uncertainty around this perceptual threshold. Humans knew when they were at risk for error in the primary discrimination, and declined those trials selectively and adaptively.

Humans' post-experimental reports constructively corroborated this primary behavioral evidence of uncertainty responding. They said that their sparse and dense responses were cued by the objective stimulus conditions (i.e., sparsity or density) on a trial. In contrast, they said that their Uncertain responses were prompted by personal feelings of uncertainty, doubt, and of not knowing the correct answer in the discrimination ("I was uncertain; I didn't know or couldn't tell"). This suggests that, for humans, the Uncertain response may reveal not only metacognitive monitoring but also a reflexive awareness of the self as cognitive monitor. Given their construal of the Uncertain response, some humans did not even like using it, for they felt it was cheating or a cop-out. Humans have given these same introspections about Uncertain responses – that is, that they are "sort of an admission of weakness" (Fernberger 1930, p. 210) – for almost 100 years. In related recent work we have found that males are especially overconfident in a task of this kind and tend to think they know even when they do not (Washburn et al. 2001). Apparently, males are unlikely to stop and ask for directions even within psychophysical discriminations.

These data replicated the phenomenon and the phenomenology of the Uncertain response in human psychophysics. They confirmed that this response is a comment by humans on the failure of the primary discriminatory process, or a no-confidence vote on that process. Humans' metacognitive performance in this task thus provided a good comparative target to which monkeys' performance could be referred.

Two 9-year-old rhesus monkeys participated in the same task. These joystick-trained monkeys were tested using the Language Research Center's Computerized Test System (LRC-CTS; Washburn & Rumbaugh 1992). They received food pellets or 19-second timeouts, respectively, for correct or incorrect responses.

Both monkeys performed like the humans did, with Uncertain responses focused on the discrimination's crossover (Figs. 3C, 3D). The monkeys declined somewhat more trials than the humans did, possibly because the monkeys (but not the humans) were working for food rewards, or possibly because the humans (but not the monkeys) had scruples about using this response. Monkeys, like humans,

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

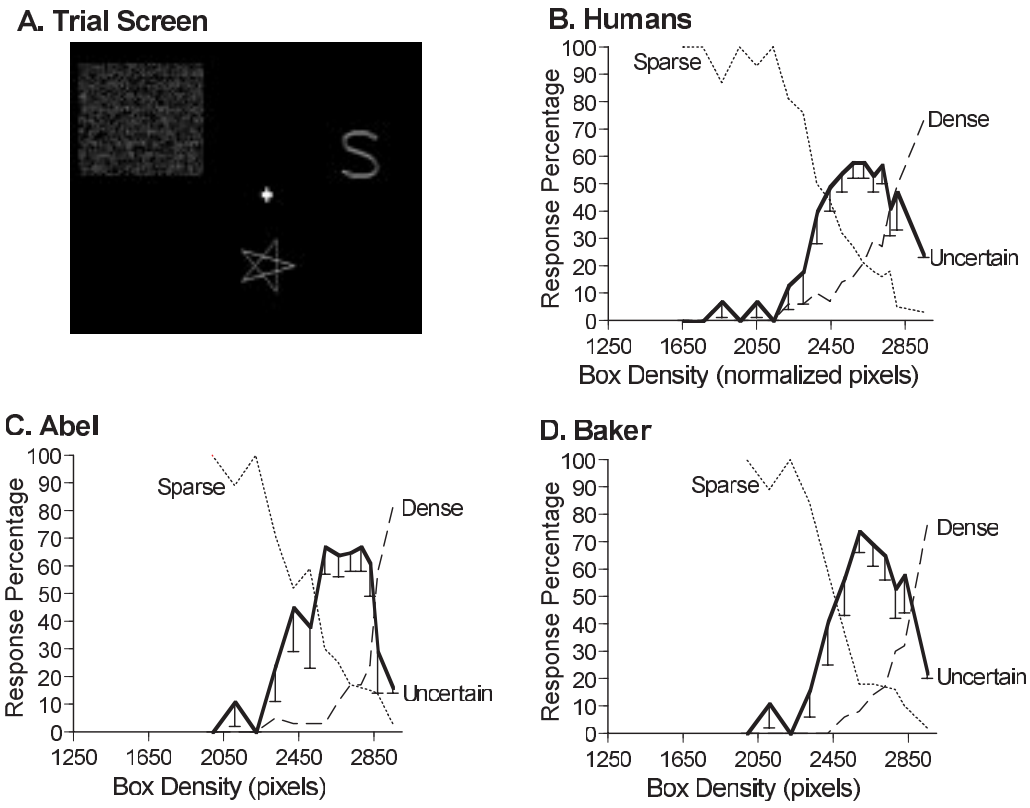


Figure 3. A. The screen from a trial in the dense-sparse discrimination of Smith et al. (1997). B. The performance of seven humans in the dense-sparse task. The dense response was correct for boxes with exactly 2,950 pixels – these trials are represented by the rightmost data point for each curve. All other boxes deserved the sparse response. To equate discrimination performance across participants, the data have been normalized to place each participant's discrimination crossover at a pixel density of about 2,700. The horizontal axis indicates the normalized pixel-density of the box. The solid line represents the percentage of trials receiving the Uncertain response at each density level. The error bars show the lower 95% confidence limits. These were calculated (Hays 1981, pp. 224–26) using the total Uncertain responses as a proportion of total trials at each density level (summing across humans who completed one session each). The percentages of trials ending with the dense response (dashed line) or sparse response (dotted line) are also shown. C. The performance of Monkey Abel in the dense-sparse discrimination depicted in the same way (here the error bars were calculated using the total Uncertain responses as a proportion of total trials at each density level summing across multiple sessions by the animal). D. The performance of Monkey Baker in the dense-sparse discrimination.

correctly assessed when they were at risk for error in the primary discrimination and declined those trials preemptively. There is clearly a strong analogy between the use of the psychophysical uncertainty response by the two species. In fact, Figure 3 presents one of the strongest existing matches between human and animal performance in the comparative literature.

A related experiment strengthened this analogy by demonstrating parallel individual differences within the two species in the use of the Uncertain response. In this experiment, instead of dynamically titrating perceivers' thresholds, we continuously offered perceivers a wide range of densities from easy sparses to easy denses. This experiment resembled others that have collected doubtful judgments from humans (e.g., Woodworth 1938).

Figures 4A and 4B, respectively, show the results from eight humans and from one particular human. The two discrimination responses (sparse and dense) were used in the same way. The difference is that the eight humans generally responded Uncertain to the difficult, indeterminate trials near the discrimination's breakpoint. The single human did not. Figures 4C and 4D, respectively, show the performance of Monkeys Baker and Abel. Both animals used the two discrimination responses in the same way, performing better for stimuli farther from the task's midpoint. Baker selectively declined the indeterminate trials near the discrimination's breakpoint that he would most probably lose; Abel did not. Thus, under the same circumstances that caused a human not to respond Uncertain in this task, Abel chose not to, either.

All humans and animals in both dense-sparse tasks just described used the primary discrimination responses (dense and sparse) in the same way. This supports the idea that the two primary responses in a discrimination task are functionally highly stable for being perceptually anchored by and mapped to the two stimulus-input classes. Across experiments, amongst humans and between monkeys, only the use of the Uncertain response showed strong differences. The early psychophysicists, observing this peculiar changeability of the Uncertain response in the hands of humans, concluded that the Uncertain response was not perceptually anchored to a stimulus input class, that it was about the failure of assignment of stimuli to an input class, and that it was related to uncertainty, extra-serial attitudes, and decisional temperaments (Angell 1907; Fernberger 1930; Thomson 1920). It is an interesting fact that monkeys'

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

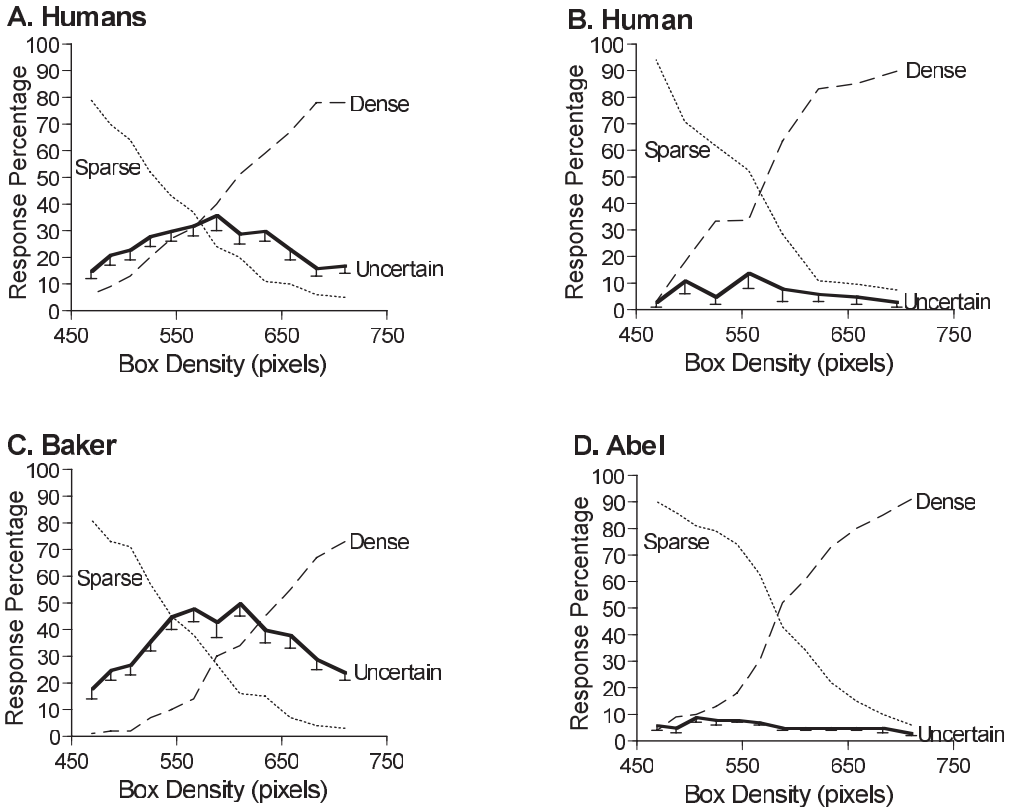


Figure 4. A. The performance of eight humans in the broad-spectrum dense-sparse discrimination of Smith et al. (1997). The upper half and lower half of a range of densities deserved the dense response and sparse response, respectively. The percentage of trials receiving the Uncertain response at each density are shown by the solid line. The error bars show the lower 95% confidence limits. The percentages of trials ending with the dense response (dashed line) or sparse response (dotted line) are also shown. B. The performance of one human in the broad-spectrum dense-sparse discrimination. C. The performance of Monkey Baker in this discrimination. D. The performance of Monkey Abel in this discrimination.

use of the Uncertain response shows this peculiar changeability, too.

7. Dolphin uncertainty responses in an auditory discrimination

Smith et al. (1995) evaluated the uncertainty-monitoring capacities of another cognitively sophisticated species by placing a dolphin in an auditory discrimination task. Pressing the high or low paddles, respectively, was correct for 2,100-Hz tones or tones of any lower frequency. The Uncertain paddle advanced the animal into an easy, low-pitched trial that was rewarded when completed with the low response. Initially, the animal was stabilized on an easy discrimination between 2,100 Hz and 1,200 Hz. Then the discrimination's difficulty was increased by raising the pitch of the below-2,100 Hz trials until the dolphin was struggling to distinguish trials of 2,100 Hz and 2,085 Hz. At mature performance, trial difficulty was adjusted based on the dolphin's performance to sustain the level of difficulty.

Figure 2 showed the dolphin's two-response performance with the Uncertain response disallowed. Low and high responses mapped to below-2,100 Hz and 2,100 Hz tones, respectively, with these response curves crossing (signifying chance performance) at the dolphin's threshold. The crucial question, not illuminated by Figure 2, was how the animal would behave at threshold when allowed to respond Uncertain.

Figure 5A answers this question. The dolphin's primary discrimination performance was the same, but now he used the Uncertain response for the difficult trials surrounding his discrimination threshold. Five humans performed similarly (Fig. 5B). Both species used the Uncertain response less in this auditory discrimination than did monkeys and humans in the density discrimination. That is, the perceivers in the auditory task had a narrower Interval of Uncertainty (Woodworth 1938). Probably the density continuum has a wider region of subjective indeterminacy because the placement of the dots in the boxes has a random element that can cloud one's perception of denseness and sparseness. Probably the pitch continuum has a narrower region of subjective indeterminacy because the pitch continuum is unidimensional and pure. It is interesting that both human and animal observers sense similarly and accurately the breadth of the zone of subjective indeterminacy and use the Uncertain response appropriately to each sensory domain.

Humans again attributed their use of the two primary discrimination responses (high and low) to the prevailing stimulus conditions (i.e., 2,100 Hz tones and lower tones). They attributed their Uncertain responses, as the early psychophysical observers did, to their states of doubt and uncertainty. Though the dolphin said nothing, an interesting

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

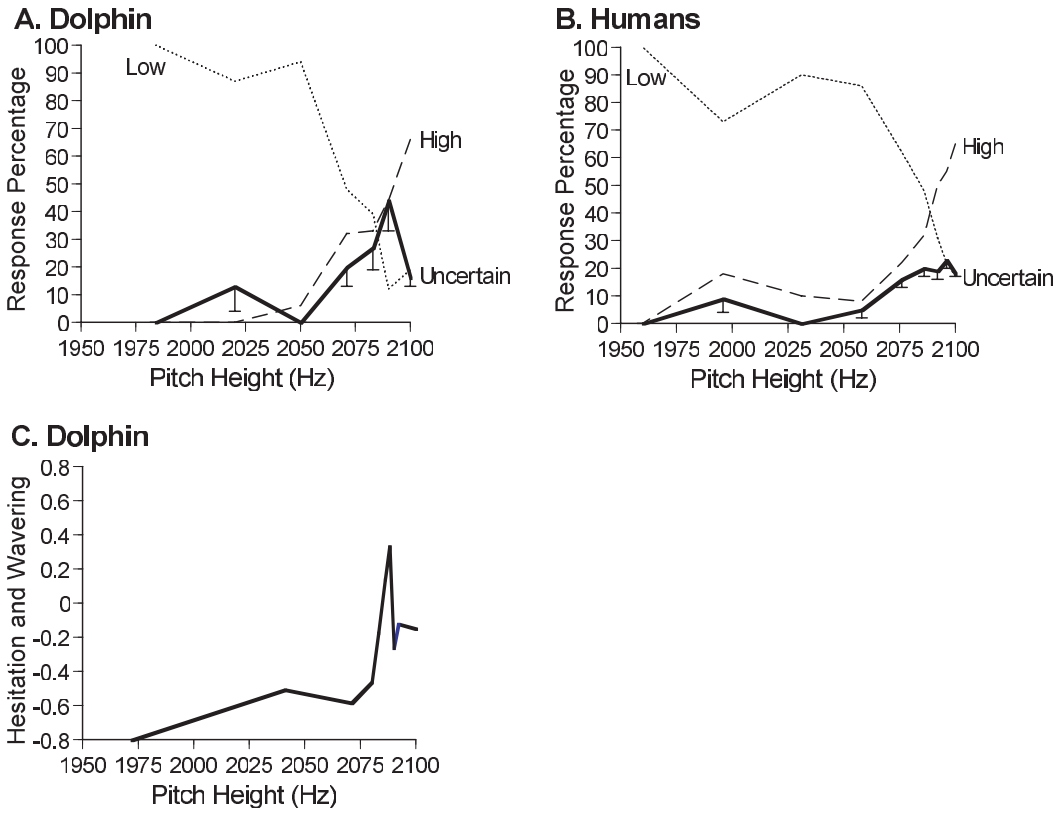


Figure 5. A. Performance by a dolphin in the auditory discrimination of Smith et al. (1995) when the Uncertain response was also available. The horizontal axis indicates the frequency (Hz) of the trial. The high response was correct for tones at 2,100 Hz – these trials are represented by the rightmost data point for each curve. All other tones deserved the low response. The solid line represents the percentage of trials receiving the Uncertain response at each pitch level. The error bars show the lower 95% confidence limits. The percentages of trials ending with the high response (dashed line) or low response (dotted line) are also shown. B. The performance of five humans in a similar auditory discrimination. C. The dolphin's weighted overall Factor 1 behavior (hesitancy, slowing, wavering) for tones of different frequencies (Hz).

additional result was that his own brand of uncertainty behaviors attended his Uncertain responses near threshold. He sometimes slowed approaching the response paddles, or wavered amongst them, or swam toward them with an open mouth, or while sweeping his head from side to side or opening and closing his mouth rhythmically. To formalize these observations, four raters judged the intensity of these behaviors during the trials in four video-taped sessions. Then factor analysis was used to discern the latent structure behind the correlations among these variables. The strongest behavioral factor was clearly allied to hesitation and wavering by the animal. Figure 5C shows the overall intensity of these Factor 1 behaviors (based on a factor-scoring procedure) for trials at different pitch levels. These behaviors peaked at 2,087 Hz and were distributed like the Uncertain response (Fig. 5A). These behaviors are the "lookings or runnings back and forth" that Tolman (1938, p. 27) thought might operationalize animal consciousness. They are also intuitive symptoms of uncertainty states in the animal. Thus, these ancillary behaviors reinforce an uncertainty interpretation of the animal's Uncertain responses. However, the Uncertain response is more easily measured and compared across situations and species than are hesitation and wavering.

8. Uncertain responses by monkeys and humans in a same-different task

The experiments just summarized focused on stimulus qualities (e.g., 2,100 Hz or lower; true dense or sparser). Those experiments leave open the possibility (discussed in sect. 14) that these Uncertain responses by animals fell under the associative control of stimulus cues rather than under the metacognitive control of uncertainty states. This possibility led Shields et al. (1997) to ask whether monkeys could recruit adaptive uncertain responses when pushed to their psychophysical limit in a same-different (SD) task. The SD task, if constructed correctly, requires a relational judgment, and an abstraction beyond the current absolute stimulus qualities, especially when sameness and difference must be judged amid variable stimulus contexts. Probably the SD task requires an additional processing step in which the relevant qualities of the two stimuli are compared to or subtracted from one another (a differencing strategy – see MacMillan & Creelman 1991). Then, as the difference resulting is near zero or larger than zero, a judgment of same or different is made. This information-processing description grounds the prominent idea that relational concepts are cognitively derived, sophisticated, and phylogenetically

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

restricted. Animals often have special difficulty with relational judgments (Carter & Werner 1978; Premack 1978; see also Cumming & Berryman 1961; Farthing & Opuda 1974; Fujita 1982; Holmes 1979) and require clever training procedures to learn them (Wright et al. 1990). Moreover, relational concepts can be fragile when placed into opposition with absolute stimulus qualities (Premack 1978). For these reasons relational learning heads standard typologies of conceptual sophistication (Herrnstein 1990). For the same reasons it is interesting to know whether animals can make adaptive uncertain responses in this abstract setting.

Accordingly, Shields et al. (1997) gave rhesus monkeys a same-different density discrimination. On each trial two rectangles filled randomly with lit pixels were shown (Fig. 6A). As the two densities shown on a trial were the same or different, animals were to make the same response (moving the cursor to the rectangles), or the different response (moving the cursor to the D). To cause animals serious difficulty, the same-different task was psychophysically scaled. That is, the size of the stimulus disparity on different trials was adjusted dynamically to challenge constantly participants' ability to discriminate same from different. In addition, same and different trials at several absolute stimulus levels were intermixed to ensure a true relational performance.

The crucial question is whether monkeys can decline trials that present indeterminate stimulus disparities. In fact they were undeterred by either the difficulty or abstractness of the task. The two monkeys (Fig. 6B) used the Uncertain response in just the way that six humans did (Fig. 6C). The animal and human performance profiles correlate at $r = 0.98$. Shields et al. (1997) even reserved some absolute density levels for transfer tasks and demonstrated that the animals were showing a true relational performance that was independent of the absolute dense and sparse stimuli that carry the relation.

9. Smith et al.'s (1998) comparative studies of memory monitoring

Memory tasks have been a sharp focus in studies of human metacognition. For example, humans can be asked to judge whether they can complete phrases like "The physicist Albert_____" or "The philosopher Albert_____." Comparative research has also asked whether animals can monitor their memory and respond adaptively when the state of their memory does not justify completing a memory test.

Smith et al.'s (1998) exploration of this capacity relied on the predictable changes in memory performance that occur across the serial positions of a memory list. The experimenter can know which items cause difficulty (and perhaps uncertainty) for the animal, and can ask whether animals

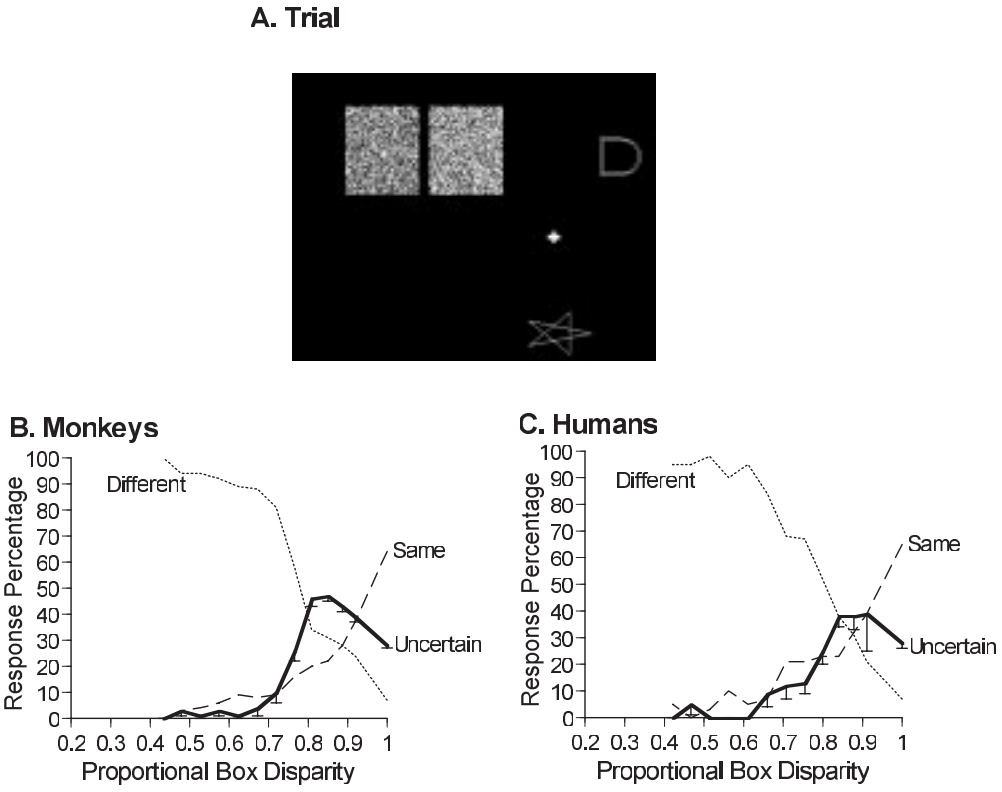


Figure 6. A. The screen from a different trial of the same-different task of Shields et al. (1997). B. Performance by two monkeys in the same-different task. The horizontal axis gives the ratio between the densities of the comparison box and the standard box for trials of different disparities. The same response was correct for trials at a proportional box disparity of 1.0, and these trials are represented by the rightmost data point for each curve. All other trials deserved the different response. The solid line represents the percentage of trials receiving the Uncertain response at each density ratio. The error bars show the lower 95% confidence limits. The percentages of trials ending with the same response (dashed line) or different response (dotted line) are also shown. C. Performance by six humans in the same-different task.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

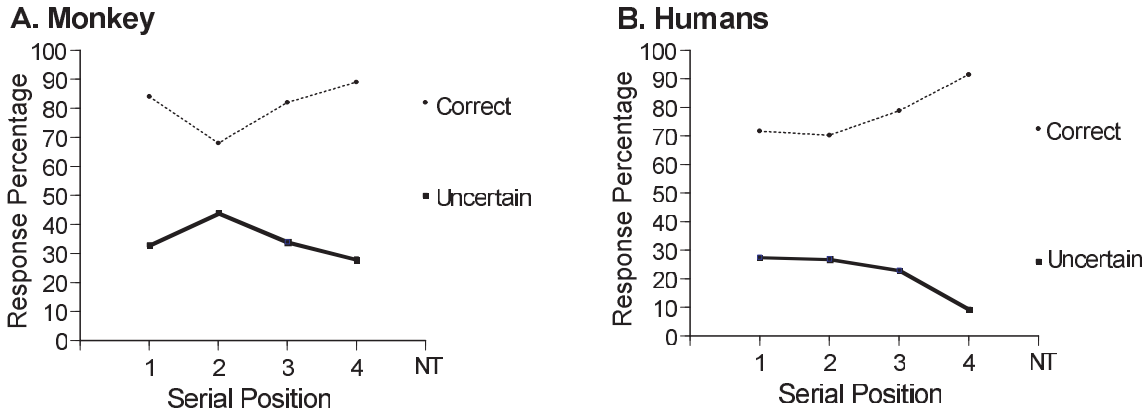


Figure 7. A. Serial probe recognition (SPR) performance by Monkey Baker in the task of Smith et al. (1998). NT denotes not there trials. The serial position (1–4) of the probe in the list of pictures is also given along the X-axis for the probes on there trials. The percentage of total trials that received the Uncertain response is shown (bold line). The percentage correct (of the trials on which the memory test was accepted) is also shown (dotted line). B. Performance by 10 humans in a similar SPR task used by Smith et al.

use the Uncertain response selectively for these difficult memory trials. Smith et al. adopted the serial-probe recognition (SPR) task that has been a staple in comparative memory research (Castro & Larsen 1992; Roberts & Kraemer 1981; Sands & Wright 1980; Wright et al. 1985). In this procedure one presents a "list" of items sequentially followed by a probe. The participant makes a there or not there response as the probe is judged to have been in the list or not. In elegant parametric research using this task, Wright et al. (1985) specified conditions that lead monkeys to show both primacy and recency effects. Our procedures followed on theirs to produce serial-position curves that had this classic shape. But we also gave animals an Uncertain response that let them decline any memory tests of their choosing.

Figure 7A shows that Monkey Baker did decline trials selectively when the middle, more difficult list positions were probed. The tendency to respond Uncertain was the mirror image of memory performance when the animal chose to complete the memory test. Figure 7B shows that the same was true for 10 humans under similar conditions (though these conditions were not suitable for humans to show a strong primacy effect). The similarity in the two graphs is especially interesting because humans were expressly instructed to use the Uncertain response as a report on memory indeterminacy. Monkeys are behaving like humans. Humans are declining memory tests when they feel uncertain.

An additional prediction follows from an uncertainty-monitoring interpretation of monkeys' performance. Animals should perform better when they *choose* to complete the memory test than when they are forced to do so. This prediction follows because the monkey should accept memory tests when he monitors quite strong or quite weak traces that can correctly be given there or not there responses. In contrast, when one forces the animal to complete all memory tests, his overall performance will be lowered by adding in the poor performance on the memory tests that he would have declined because he monitored indeterminate traces on those trials.

To evaluate this important prediction, Smith et al. (1998) also ran Monkey Baker without the Uncertain response available. Under these conditions, Baker was 72%, 51%, 67%, and 75% correct, respectively, when serial positions 1 to 4 were probed. His performance was 12%, 17%, 15%, and 14% higher, respectively, when he had the Uncertain response available but chose to complete the memory test. Baker increased his rewards per minute by 18% by using the Uncertain response adaptively to avoid errors when he monitored an indeterminately available memory. Baker was clearly sensing something real about his memory that was rationally attended to in deciding to accept or decline memory tests. This prediction – of a performance advantage on chosen trials over forced trials – figures prominently in a variety of related studies, the next of which we consider now.

10. Hampton's (2001) comparative studies of memory monitoring

Hampton also asked whether monkeys can use the monitored strengths of memory traces in deciding whether to accept or decline memory tests. Hampton's exploration of this capacity took advantage of the fact that memory performance declines predictably during a forgetting interval. Thus, the experimenter can know which items cause difficulty (and perhaps uncertainty) for animals, and can ask whether animals use the Uncertain or Decline response selectively for these difficult memory tests. Hampton adopted the delayed matching-to-sample (DMTS) task that has also been a staple in comparative memory research (Elliott & Dolan 1999; Etkin & D'Amato 1969; Herman 1975). In this procedure one presents a sample that then must be chosen after the forgetting interval from among two or more alternatives.

The critical point in Hampton's procedure came at the end of the forgetting interval on each trial. Then, on 67% of trials, the animal chose, by making one or another discriminative response, either to accept the memory test (a four-alternative, forced-choice test in which the sample was presented along with three foils) or to decline it. Accepted memory tests led to either a preferred food reward or a timeout for correct and incorrect responses, respectively. Declined memory tests led to a non-preferred food reward with no risk of timeout. The ideal strategy was to consult the strength or availability of the sample's trace in memory at

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

the end of the forgetting interval, and accept the trial if that strength exceeded some criterion value that would probably support correct recognition in the memory test. By this strategy, the animal would sensibly decline tests when the trace fell below criterion because that weak trace might be unrecognizable amidst the foils at test. On the remaining 33% of trials, the animal was forced to complete the memory test, just as Monkey Baker was when Smith et al. (1998) ran him without the Uncertain response available. Note that by interspersing chosen and forced memory tests, Hampton was able to simultaneously monitor the animals' performance on these two kinds of trials. This is an innovative approach that both Teller (1989) and Inman and Shettleworth (1999) developed independently. Hampton predicted that forced memory tests would produce poorer performance than chosen memory tests for the reason already given.

Hampton's Experiment 3 is especially interesting and critical. The forgetting interval was varied in several steps from about 15 sec to more than 100 sec. Hampton made two predictions that would follow if animals were accepting or declining memory tests based on metamemory. First, they should decline more trials as the forgetting interval increased, because the sample's trace would grow less available with time and would more often fall below the criterion level of strength or availability. Second, animals should show a stronger advantage in performance for chosen memory tests over forced memory tests as the forgetting interval increased. This would occur because the strong traces after short delays would support near-ceiling performance on chosen and forced trials, whereas the weaker traces after long delays would be more variable and, if the animal monitored the strength of these traces, would show more clearly the adaptive value of the metacognitive strategy.

One monkey confirmed both predictions (Table 1, Row 1), showing a perfect metamemory data pattern. In the table, DL1 to DL4 denote four increasing difficulty levels, which here refer to longer forgetting intervals. (Hampton did not fully report performance for a fifth, longest delay. The performance estimates in that condition were unstable because nearly all trials were declined. So we did not include this fifth delay condition here.) More memory tests were declined at longer delays. There was better performance on chosen than on forced memory tests, especially at long delays. The second monkey (Table 1, Row 2) presented a different data pattern. This animal did decline more memory tests at longer delays. Alone, though, this result might simply reflect that the animal associated long delays with memory-test errors and timeouts, motivating decline responses after long delays. This animal hardly showed the other crucial component of the metamemory data pattern (performing better on chosen than on forced trials).

The metamemory studies of Smith et al. and Hampton are similar but have an important difference and complementarity. The difference is that Hampton's monkeys needed to place only one criterion along the continuum of trace-strength impressions. They decided only whether to accept the memory test or decline it. Smith et al.'s monkeys needed to place two criterion lines because if the trace were really available or unavailable they needed to respond "there" or "not there," respectively, so that they only declined the memory test for indeterminate traces. We will return to the

Table 1. *Percentage of chosen and forced memory tests answered correctly, and percentage of memory tests declined, for four levels of trial difficulty (DL1 to DL4) in previous studies*

	Percentage Correct on Chosen Tests				Percentage Correct on Forced Tests				Percentage of Memory Tests Declined			
Source	DL1	DL2	DL3	DL4	DL1	DL2	DL3	DL4	DL1	DL2	DL3	DL4
1. Hampton (2001) Monkey 1	94	90	90	78	92	88	65	52	20	18	52	80
2. Hampton (2001) Monkey 2	93	91	79	55	93	89	75	50	10	23	35	48
3. Monkey 1 (simulated STM strategy)	97	97	92	91	92	87	65	50	14	18	55	74
4. Monkey 1 (simulated LTM strategy)	96	93	80	69	91	89	64	52	15	23	56	71
5. Monkey 2 (simulated)	94	89	76	54	93	88	74	50	15	19	31	52
6. Shields (1999) Monkey (prospective)	78	73	56	30	77	71	55	30	44	45	46	46
7. Shields (1999) Monkey (simultaneous)	83	76	62	36	77	69	54	32	37	45	57	58
8. Teller (1989) Pigeons	72			37	70			36	40			60
9. Inman/Shettleworth (1999) Pigeons (E1)	90	84	79	74	86	77	75	68	33	35	38	40
10. Inman/Shettleworth (1999) Pigeons (E2)	90	90	80	76	88	83	81	73	41	41	48	46

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

question of how the single criterion line in Hampton's task relates to the criterion lines in Smith et al.'s task.

The complementarity is that Smith et al. and Hampton followed different lines of research in the human metacognition literature. Smith et al. asked their animals for the decline-accept decision with the memory probe present. The animal had to consider how available that probe's trace was in the context of the memory list – that is, whether it was easily available as a member of the list, or unavailable, or indeterminate. Hampton asked his animals for the decline-accept decision with the memory probe absent. The animal had to consider how active relevant traces were in memory and whether they were active or accessible enough to suggest accepting the memory test. Hampton's procedure is a strong one, because the animal was not shown which memory location should be evaluated in making the decline-accept decision. (Inman & Shettleworth 1999 also used this procedure in their Experiment 2, which is discussed in sect. 12.3). The limitation on Hampton's method is that only four pictures were ever relevant at a time, so the animal only needed at most to consult the availability of all four traces. Smith et al.'s procedure is a strong one because it is known that re-presenting the material to be judged for the quality of memory remakes those traces active and available and makes metacognition judgments more difficult (Dunlosky & Nelson 1992; 1997). The limitation to this approach is that Smith et al. showed their animals which memory location to monitor for a contextual list memory. In the section on formal approaches we will see that Smith et al.'s and Hampton's procedures are analyzable using the same formal assumptions, suggesting that both studies tapped the same memory-monitoring capacity in monkeys. However, as we will see next, neither procedure meets the full challenge posed by the human paradigms. We describe now an experiment that came closer to doing so, though the monkeys did not meet the challenge because they did not express a metacognitive capacity within it.

11. Shields's (1999) comparative studies of memory monitoring

Shields (1999) undertook the comparative study of memory monitoring that is closest to the human paradigms. This study suggests limits on monkeys' metamemory and indicates lines of future research. Human metamemory experiments frequently ask participants to make feeling-of-knowing judgments about linked information that is not presented, so that the query prompts a search for the availability of information at some other memory location (e.g., _____Putin). In contrast, both Smith et al. (1998) and Hampton (2001) asked directly about the availability of memory material that itself had just been presented (e.g., Vladimir – 30 sec forgetting interval – Igor or Vladimir or Ivan?). Accordingly, Shields trained animals in a paired-associate (PA) task in which links were established in memory between arbitrary sample-target pairs of nine-point polygon shapes. Using shapes dodged the limitation that animals do not know about politicians. On one screen of the task, animals saw the sample and the Star. If they chose the sample (a judgment of knowing), they were tested earning food rewards or long timeouts on the subsequent screen by having to choose between the associated target for the sample and a foil (another sample's target). If they chose the Star (a judgment of not knowing), they were tested on the subsequent screen with the possibility of reward but no risk of a long timeout. Testing the animals even after they responded Uncertain mirrored the important aspect of human metamemory experiments that humans attempt recognition both after feelings of knowing and after feelings of not knowing. Comparing these two performance levels lets one confirm that participants are uncertain when they respond Uncertain. This feature of this experiment was balanced against the risk that this would make the function of the Uncertain response more difficult for animals to grasp or its use too attractive. The idea in the experiment was that the sample would be the query that prompted the animal to judge whether it knew the target (i.e., the linked memory material). Therefore, a critical feature of Shields's task was that sample-target pairs occurred at different repetition rates so that some became better learned than others.

Table 1 (Row 6) shows a monkey's performance in this procedure. The sample-target pairs occurring at different repetition rates are treated as Difficulty Levels 1 to 4. The monkey learned more poorly the sample-target pairs that occurred more rarely. But he was unable to decline those trials selectively based on seeing the sample alone. Nor was there any difference in performance between the trials he chose to accept or to decline. Shields measured performance on chosen and declined trials separately, whereas Hampton measured performance on chosen and forced trials (the latter a combination of trials that would have been chosen and declined). To make Table 1 uniform throughout, we combined the chosen and declined performance levels algebraically into an estimate of the monkey's performance had he been forced to complete some memory tests. Thus we set Forced Percent Correct = Proportion Chosen $\times$ Percent Correct Chosen + Proportion Declined $\times$ Percent Correct Declined.

Thus, the monkey showed no evidence of making Uncertain responses that were based in metamemory. Given this failure, Shields (1999) adjusted the procedure. She let the second test screen contain the sample at the top, the target and foil below it, and the Star below them. Now the monkey had visibly present all the information he needed to judge whether he should accept the memory test. This time he did respond Uncertain more often to less-well-known sample-target pairs (see Row 7), and he did show a performance advantage when he chose to accept the memory test. For some reason, though, the complexity of the PA task made it difficult for him to show this data pattern based on seeing the sample alone. It remains a research challenge to evaluate whether this is a real limit on animals' monitoring capacities, or whether something in Shields's procedure produced it. In any case, Shields's experiment suggests constructive ways to bring animal and human metamemory assessments closer together.

12. Tests of uncertainty monitoring in less cognitively sophisticated species

This result, suggestive of limits on monkeys' metacognitive capacities, in turn makes one wonder about the monitoring capacities and limits of other species. Several researchers have looked for metacognitive capacities in species such as rats and pigeons that are less cognitively sophisticated and more associative in their behavioral solutions. Some pre-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

liminary comments are worthwhile here. The available studies do not provide strong evidence for metacognitive capacities in these species. The negative data patterns could be important theoretically for suggesting that these species are not self-reflective in the way required to monitor uncertainty states and to respond adaptively to them. The phylogenetic map showing the emergence of self-reflective cognition in some species and not others would be exciting to see. However, one must also remember to interpret the null findings cautiously, taking into consideration whether the experimenters did enough of the right things for long enough to create conditions that ought to have let animals show a metacognitive capacity if they have it. This is a high hurdle for an experimenter to meet, because more extensive testing, or different motivators, and so forth, might have revealed the metacognitive capacity when the actual procedures did not.

Our view is that the three studies that we discuss below did a good job of arranging circumstances appropriately, so that the negative results provided in these three studies may be interpretable and important. However, readers will wish to make their own judgment on this point.

12.1. Uncertain responses by rats in an auditory-discrimination task (Smith & Schull 1989)

Smith and Schull (1989) adapted the psychophysical approach to studying uncertainty monitoring (Smith et al. 1995; 1997) by placing rats in a pitch-discrimination task. Animals were to make a left or right lever press, respectively, when they heard a repeating 400-Hz tone or a 400-Hz tone alternating with any other. A third response let the animals decline a trial at any time and begin a new trial instead. Animals knew well the effect of this response because they also used it to initiate trials. At first, participants were stabilized on an easy discrimination between 400–400 Hz tone pairs and 400–700 Hz tone pairs. Then the difficulty of the discrimination was increased by decreasing the frequency of the higher pitch until performance faltered at tone pairs of about 400–410 Hz.

Figure 8 summarizes the performance of six rats. On alternating trials, the alternating response predominated, whereas the repeating response predominated on repeating trials and on the most difficult alternating trials. Where these two response curves cross, the primary discrimination was performed at chance. The rats mastered well the two primary discrimination responses. But they did not recruit the Uncertain response adaptively at their threshold, as humans, monkeys, and a dolphin did. The rats did not assess when they were at risk for error in the primary discrimination and decline those trials selectively and adaptively.

We wondered whether the animals' problem was that the Uncertain response doubled as a trial-initiation response. To check on this, we incorporated an additional trial type (a higher tone repeating faster) that was rewarded randomly so that animals could only be 50% correct on this trial type (just as they are at threshold). The rats responded Uncertain to these trials three times as often as they responded Uncertain at threshold. This shows that they were associating some trials with lean reinforcement and that they were prepared to wave off these trials. In fact, it is interesting that animals declined the objective, stimulus-borne 50:50 contingency, but not the subjective, threshold-borne 50:50 contingency. This could suggest that the latter, subjective

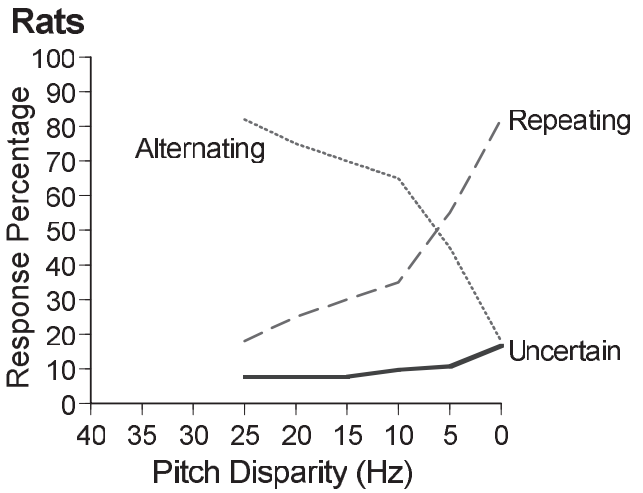


Figure 8. Performance by six rats in the frequency discrimination used by Smith and Schull (1989). The horizontal axis indicates the frequency difference between the alternating pitches on a trial. The repeating response was correct for trials with a frequency difference of 0, and these trials are represented by the rightmost data point for each curve. All other trials deserved the alternating response. The solid line represents the percentage of trials receiving the Uncertain response at each pitch disparity. The percentages of trials ending with the repeating response (dashed line) or alternating response (dotted line) are also shown.

cue runs more quietly and less accessibly in the rat's cognitive system, whereas it seems lively and accessible in the cognitive systems of humans, monkeys, and dolphins.

We do not conclude from this that rats cannot monitor their uncertainty and decline trials based on it. Under different circumstances or methodology, they might (e.g., if one combined a task in their strongest sensory modality with contingencies that strongly favored the use of the Uncertain response). However, Smith and Schull's version of the experiment continued for months while every day we tried unsuccessfully to coax the monitoring capacity from these animals. Perhaps it is difficult to document a sensitivity to these uncertainty or metacognitive cues because they are difficult for the rat to sense and use. This same difficulty arises regarding pigeons' cognitive systems, too.

12.2. Evaluating pigeons' capacity for metamemory in a DMTS task (Teller 1989)

Teller (1989) was the first to study metamemory in the pigeon. As an undergraduate thesis, this project deserves special mention for its original contribution. Its methodology foreshadows strikingly the one used in Hampton (2001) and Inman and Shettleworth (1999 – to be described next). Six animals participated in a DMTS task with trials containing either 0sec or about 28sec forgetting intervals. There were two reinforcement schedules. By one discriminative response, the bird chose to complete the memory test with no hint information (i.e., without the sample highlighted on the screen as the obvious correct answer), receiving either no reward or a large grain reward. By another discriminative response, the bird chose to complete the memory test with the hint provided, almost surely receiving a small grain reward. On 60% of trials, birds chose between these schedules. When they chose the no-hint schedule, their performance was comparable to performance on Hampton's cho-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

sen trials. On 40% of trials, the animals were required to operate under either the hint or no-hint schedules. The forced-hint trial type balances the design but is uninteresting. The forced-no hint trial type is important. Performance on these trials is comparable to performance on Hampton's forced trials.

Under the hypothesis of metamemory, pigeons should choose the hint schedule (i.e., decline the memory test) more often after the delay. And they should perform better when they choose to complete the memory test than when they are forced to. The first prediction was confirmed (Table 1, Row 8 – the 0sec and 28sec delays are treated as Difficulty Levels 1 and 4, respectively). Remember, though, that this could reflect only that the choice of schedule has come under the associative control of the delay length. The birds were not more accurate when they chose the no-hint schedule than when they were forced to use it, as they should be if their choice was based on the monitoring of a more available memory trace. Their performance was similar to that of Hampton's Monkey 2 (Row 2), who also showed a minimally metacognitive data pattern. Teller concluded that his results did not demonstrate a capacity for memory monitoring in pigeons, though he suggested that future research might be able to do so and focused his discussion on methodological considerations regarding future research.

12.3. Evaluating pigeons' capacity for metamemory in a DMTS task (Inman & Shettleworth 1999)

Inman and Shettleworth (1999) evaluated pigeons' metamemory capacity using a similar but independent approach. In Experiment 1, four birds saw one of three possible samples followed by a forgetting interval of 1–8sec. On 33% of trials, a normal DMTS memory test (that could earn a large food reward) followed the delay. These trials, on which the animal was forced to complete the memory test, correspond to the forced trials in Teller (1989) and in Hampton (2001). On 33% of trials, the delay ended with a safe response that earned a small food reward. As in Teller's case, these trials balance the design and are uninteresting. On 33% of trials, the delay ended with the safe response and the DMTS memory test presented in combination, so the animal could choose whether to decline or accept the test. The trials accepted of these combined trials correspond to the chosen trials in Teller and Hampton. The familiar predictions from this experiment (under the metamemory hypothesis) are: (1) the use of the safe key to decline memory tests should increase at longer delays; and (2) there should be a performance advantage for chosen over forced memory tests, especially at longer delays, as one of Hampton's monkeys showed but as Teller's pigeons did not show.

Inman and Shettleworth observed (Table 1, Row 9) that longer delays only produced a 7% increase in the use of the safe response (not significant by a parametric test), even though the birds performed worse at the longer delays and should have used that response more (as Hampton's and Teller's animals did). Moreover, there was no reliable performance advantage for chosen trials over forced trials, as one of Hampton's monkeys clearly showed, and there was no interaction between the length of the delay and the size of this advantage, as should be true under a metacognitive hypothesis.

In Experiment 2, Inman and Shettleworth asked four pigeons to choose between accepting the memory test and declining it (via the safe response) after the forgetting interval but before the choice objects were revealed. This experiment is almost identical to Hampton's monkey experiment, but with different results (Table 1, Row 10). Again longer delays only produced a small, nonsignificant increase (5–7%) in the use of the safe response. Again there was no advantage for chosen trials over forced trials, and there was no interaction between the length of the delay and the size of this advantage.

Inman and Shettleworth concluded appropriately that their data did not show that pigeons used memory-trace strength as a discriminative stimulus, though they acknowledged it remained possible birds could do so (e.g., if a wider range of matching accuracies was sampled, perhaps by increasing forgetting intervals beyond 8sec). Like Smith and Schull with rats and Teller with pigeons, Inman and Shettleworth suggested that pigeons might have only a weak metamemory capacity, making it difficult for the experimenter to observe it because it is difficult for the birds to use.

13. A unifying formal perspective

Now we offer a unifying formal description of performance in the tasks reviewed here. This description serves several useful ends. It supports cross-task and cross-species comparisons among data patterns. It allows the studies of optimality that figure prominently in discussions of animal behavior. It offers a neutral description of performance that is inclusive theoretically because it makes no theoretical commitments toward behaviorism or cognitivism. It clarifies the formal structure of behavior so that different theoretical perspectives can be brought to bear on it.

Our model is grounded in Signal Detection Theory (SDT – MacMillan & Creelman 1991). SDT assumes that performance in perceptual or memory tasks is organized along an ordered series (a continuum) of psychological representations of changing impact or increasing strength. In the same-different task, for example, the continuum of subjective impressions would run from clearly different (a large disparity between stimuli) to same (zero disparity – the X-axis in Fig. 9A). Given this continuum, SDT assumes that an objective event will create subjective impressions from time to time that vary around some average impression. A threshold different trial might create impressions that vary as Figure 9A's D (Different) normal distribution. Same trials might create impressions that vary as Figure 9A's S (Same) normal distribution. The overlap between these distributions is what ensures errors and fosters uncertainty in the task – both kinds of trials can feel alike to the perceiver. Finally, SDT assumes a decisional process by which criterion lines are placed along the continuum so that response regions are organized. In Figure 9A, as stimulus pairs made disparity impressions that fell to the left of the Different-Uncertain criterion line, to the right of the Uncertain-Same criterion line, or between these two, the participant would make the Different (D), Same (S), or Uncertain (U) response, respectively.

We will illustrate the application of the SDT model to animals' performances in the same-different (SD) task of Shields et al. (1997), the serial probe recognition (SPR) task

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

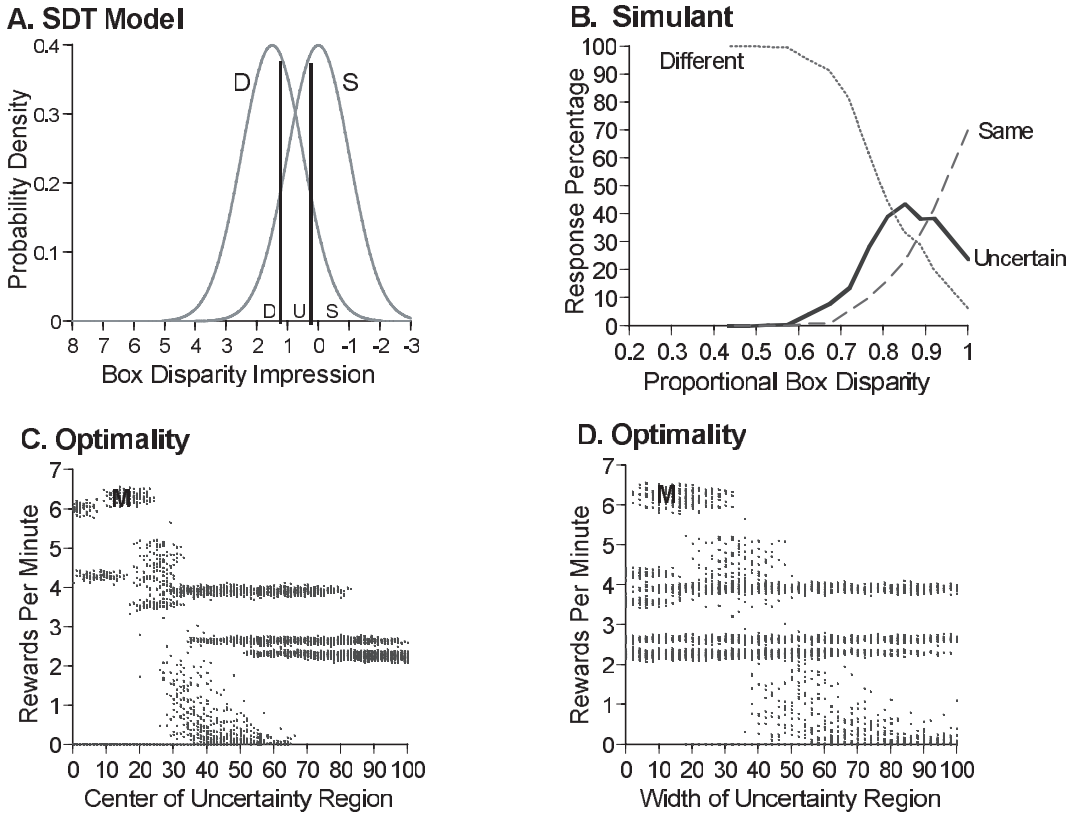


Figure 9. A. A signal detection theory (SDT) portrayal of monkeys' decisional strategy in the same-different task of Shields et al. (1997). Unit-normal disparity-impression distributions are centered at 0.0 for Same (S) trials and at a positive disparity for Different (D) trials. These normal curves are overlain by decision criteria that define the animal's three response regions (from left to right, Different [D], Uncertain [U], and Same [S]). B. Performance by the simulant that fit best the monkeys' performance (compare Fig. 6b) in the same-different discrimination task of Shields et al. (1997). The horizontal axis gives the ratio between the densities of the comparison box and the standard box for trials of different disparities. The same response was correct for trials at a proportional box disparity of 1.0 – these trials are represented by the rightmost data point for each curve. All other trials deserved the different response. The solid line represents the percentage of trials receiving the Uncertain response at each density ratio. The percentages of trials ending with the same response (dashed line) or different response (dotted line) are also shown. C. The reward efficiency (in rewards earned per minute) of simulants that centered the Uncertain response region at different places along the disparity-impression continuum in a virtual version of Shields et al.'s (1997) same-different discrimination. We surveyed the reward efficiency of 5,151 decisional strategies when each received 8,000 trials in a simulated version of the task, subject to the trial times, penalty times, and reward structure of the task the monkeys experienced, and using the signal-detection response rule that accorded with the three response regions defined by each simulant's two criterion placements. M represents the position in this optimality space of the simulant that best fit the performance of the real monkeys. D. The results of the same simulation plotted by the width of the Uncertain response region.

of Smith et al. (1998), and the delayed matching-to-sample (DMTS) task of Hampton (2001). The details of the simulations are given in Appendix 1.

13.1. The SD task of Shields et al. (1997)

The goal of applying the SDT model was to find the performance parameters of the simulated perceiver who – while conforming to the model – produced performance most like that which the monkeys showed (Fig. 6B). The crucial step in modeling was to assess the decisional strategy that monkeys probably used by sampling many different placements for the Different-Uncertain and Uncertain-Same criteria shown in Figure 9A. In fact, we modeled the performance of 520,251 simulated perceivers (hereafter simulants) who had different decision criteria. Figure 9B shows the performance of the simulant among these whose performance most closely matched that of the monkeys (compare Fig. 6B). The predicted response percentages were within 3–4% of their observed targets. This fit compares favorably with the fit of other formal models in the experimental literature (e.g., Smith & Minda 1998; 2000). The best-fitting simulant placed its criteria at .825 (Different-Uncertain) and .905 (Uncertain-Same) along the subjective-disparity continuum.

The SDT model also lets us assess the optimality of the monkeys' decisional strategy. To do so, and to illustrate methods in this area, we evaluated the reward efficiency of 5,151 simulants who performed the SD task using variously placed decision criteria while also experiencing virtually the trial times and punishment times animals experienced. That is, we calculated the rewards per minute that each decisional strategy would earn. Figures 9C and 9D show the rewards per minute earned by simulants that centered their Uncertain response region at different places along the disparity continuum and that gave this region different widths. "M" denotes the monkeys' best-fitting simulant. The monkeys' decisional strategies were essentially optimal because

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

they centered their Uncertain response region at their threshold for discriminating same from different and because they widened it judiciously, too.

We point out that this optimality study, like the SDT model it is based on, is psychologically neutral regarding the processes and representations that underlie performance and regarding the level in the cognitive system these processes and representations occupy. Animals could regulate optimally using low-level, high-level, or even conscious processes. Animals could regulate poorly at these levels, too. It is a step beyond finding animals' positions in an optimality landscape to judge the cognitive sophistication of the decisional strategy that placed them there.

13.2. The SPR task of Smith et al. (1998)

In this case, SDT would assume that the list items create subjective memory impressions that lie along a continuum of trace strength (the X-axis in Fig. 10A). The probe then queries the strength of one trace. Probes on Not There trials will generally point to weak traces, perhaps averaging 0.0 plus or minus the scatter of memory variability (the normal distribution NT in Fig. 10A). Probes on There trials will point to stronger traces on average (though still with memory variability), especially for the primacy and recency list items (the four T normal distributions in the figure). The overlap between the Not There and There distributions is what makes the SPR task difficult and uncertain.

Once again SDT assumes a decision process by which criterion lines are placed along the continuum to define response regions. In Figure 10A, as a probe stimulus contacted a trace that fell to the left of the Not There-Uncertain criterion line, to the right of the Uncertain-There criterion line, or between these two, the participant would make the Not There, There, or Uncertain response, respectively.

To find the best-fitting configuration of the SDT model, we modeled the performance of 226,981 simulants with differently placed decision criteria. Figure 10B shows that the

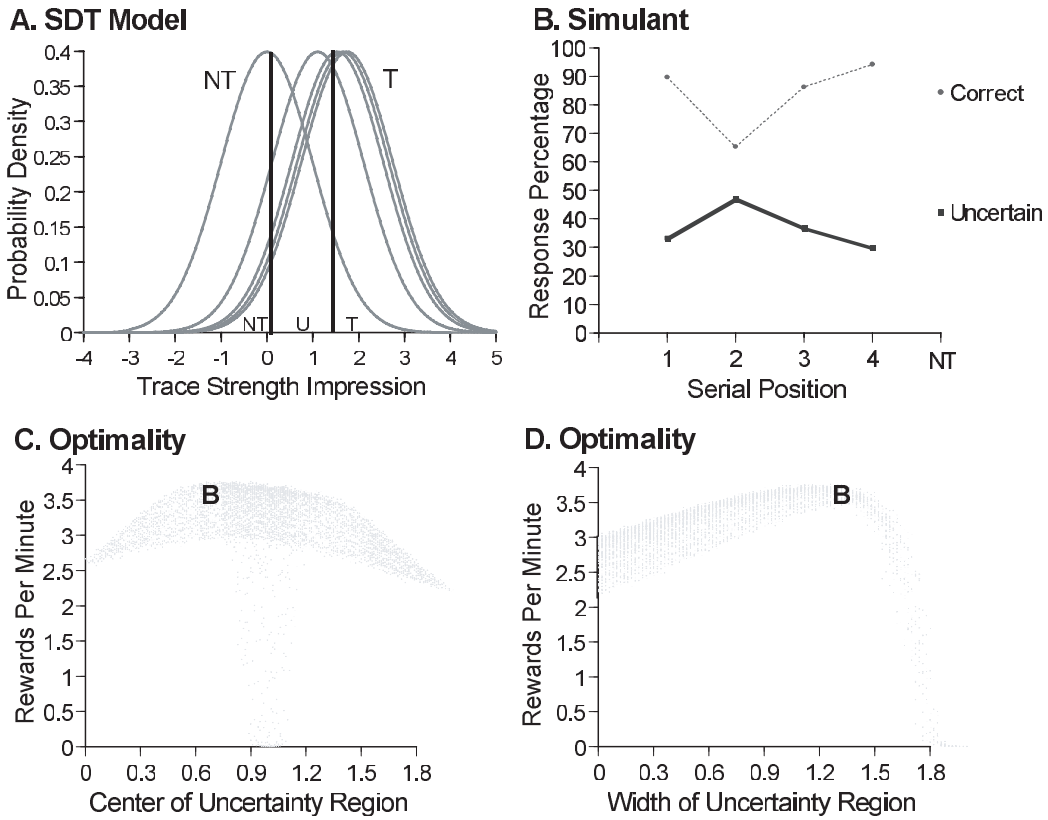


Figure 10. A. A signal detection theory (SDT) portrayal of Monkey Baker's decisional strategy in the serial probe recognition task of Smith et al. (1998). Unit-normal trace-impression distributions are centered at the locations along the trace-strength continuum corresponding to the animal's *d'* for probes of the four serial positions in the memory lists (T), and at 0.0 for the not there probes (NT). These normal curves are overlain by the decision criteria that define the animal's three response regions (from left to right, Not There [NT], Uncertain [U], and There [T]). B. Performance by the simulant that fit best Monkey Baker's performance (compare Fig. 7a) in the serial probe recognition task of Smith et al. (1998). NT denotes Not There trials. The serial position (1–4) of the probe in the list of pictures is also given along the X-axis for the probes on There trials. The percentage of total trials receiving the Uncertain response is shown (solid line). The percentage correct (of the trials on which the memory test was accepted) is also shown (dotted line). C. The reward efficiency (in rewards earned per minute) of simulants that centered the Uncertain response region at different places along the trace-strength continuum in a virtual version of Smith et al.'s (1998) serial probe recognition task. We surveyed the reward efficiency of 5,151 decisional strategies when each received 8,000 trials in a simulated version of the task, subject to the trial times, penalty times, and reward structure of the task the monkeys experienced, and using the signal-detection response rule that accorded with the three response regions defined by each simulant's two criterion placements. B represents the position in this optimality space of the simulant that best fit the performance of the real Monkey Baker. D. The results of the same simulation plotted by the width of the Uncertain response region.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

performance of the best-fitting simulant closely reproduced Baker's performance (compare Fig. 7A). On average, the response percentages were within 3% of their observed targets. Figure 10A shows the SDT description of this simulant.

We also drew the optimality landscape of Smith et al.'s SPR task. To do so, we evaluated the reward efficiency of 5,151 simulants that centered the Uncertain response region at various places along the trace-strength continuum and that widened it to varying degrees. Figures 10C and 10D show the rewards per minute of simulants that gave the Uncertain response region different centers and widths. "B" indicates the position in these landscapes of Monkey Baker's best-fitting simulant. Here, too, the monkeys centered and widened their Uncertain response region adeptly. As Figure 10A shows, they declined those trace strengths that were most indeterminate and that most risked error.

Optimality studies like these can support one's experimental planning in this area. They let one preview how different penalties and rewards change the shape of the optimality surface, and they let one find experimental settings that emphasize the value of the metacognitive strategy over alternatives. This may encourage animal participants to adopt the metacognitive strategy if they can. This preview may be especially important in the domain of comparative metacognition for this reason. We believe that the metacognitive strategy is a subtle and effortful one even for monkeys. In our experience, even monkeys gravitate toward an associative, nonmetacognitive performance strategy if they can find an effortless one that earns a decent rate of return. Hampton's Monkey 2 may illustrate this tendency. An experiment that creates the maximal separation between the metacognitive and nonmetacognitive strategies on the optimality surface may help convince monkeys that the cognitive effort of the former is worthwhile. For rats and pigeons, this maximal separation favoring metacognitive monitoring may be even more critical because these species have difficulty expressing the metacognitive capacity at all.

13.3. The DMTS task of Hampton (2001)

Regarding Hampton's Monkey 1, we assumed (Fig. 11) that samples left behind strong or weak impressions after short or long delays (the four S [Sample] normal distributions), and that foils (non-sample items) were represented by weaker memory impressions that averaged 0 with memory variability (the F [Foil] normal distribution). We also assumed that the animal placed one criterion line on this trace-strength continuum to separate the Decline and Accept response regions. In this way he would decline or accept memory tests as the monitored trace was weaker or stronger than this Decline-Accept criterion. Notice how the crucial predictions from memory monitoring emerge, given this model. First, the lower trace strengths given longer delays will mean more below-criterion sample traces and thus higher percentages of trials declined. Second, if the animal evaluates the availability of the sample's trace in memory, then he can selectively choose to accept memory tests when he monitors strong sample traces and he will perform especially well on these trials he accepts.

To find the best-fitting parameter configuration of the SDT model, we modeled the performance of 201 simulants that had differently placed Decline-Accept criteria. Table 1 (Row 3) shows that the best-fitting simulant reproduced

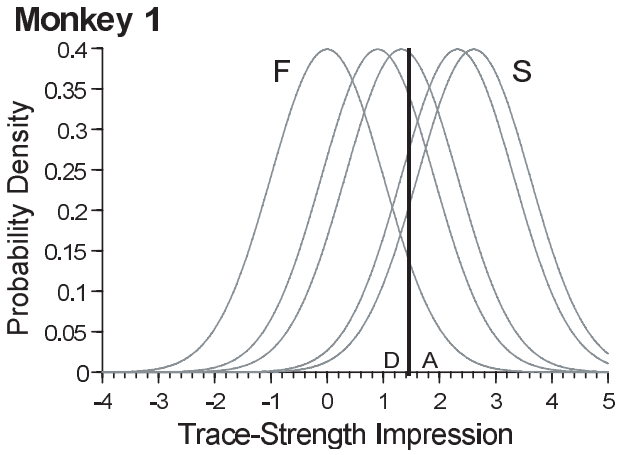


Figure 11. An SDT portrayal of Monkey 1's decisional strategy in the delayed matching-to-sample task of Hampton (2001). Unit-normal trace-impression distributions are centered at the locations along the trace-strength continuum corresponding to the animal's *d'* for samples remembered after four forgetting intervals (S), and at 0.0 for foils not seen on that trial (F). These normal curves are overlain by the decision criterion that defines the animal's two response regions (from left to right, Decline [D] and Accept [A]).

well the observed performance (Row 1). The predicted response percentages were within 3–4% of their observed targets. Figure 11 shows an SDT description of this simulant.

It is worth comparing the SDT descriptions of the memory-monitoring performances achieved by Baker and by Monkey 1 (see Figs. 10A and 11). Monkey 1, of course, had no Not There response region and so lacks the second, lower criterion point. But the decision he had to make (whether the monitored trace was strong enough to accept the memory test) is like the second decision that Monkey Baker had to make (whether the monitored trace was strong enough to respond There). It is interesting that across laboratories, methodologies, and monkeys this upper criterion lies at the same place on the trace-strength continuum. Further work could pursue similarities like these and possibly establish the point of confidence at which animals think they know or remember. Meanwhile, this similarity suggests that Baker and Monkey 1 were responding in a similar way to the same memory-strength cue.

13.4. Formal modeling and theory development

Our hope is that formal models like these could eventually support theory development in this area. We illustrate the kind of theoretical insight they might offer. There are two memory-monitoring strategies that Monkey 1 could have used in Hampton's experiment. By one strategy, based in short-term memory, the animal would "rehearse" the sample as best he could, and after the delay assess whether that one trace was still available enough to accept the memory test. By another strategy, based in longer-term memory, the animal would passively wait out the delay and then query all four relevant memory locations to see whether any trace was available enough to accept the memory test (for Hampton's monkeys, only four stimuli were relevant at a time).

Which of these strategies was Monkey 1 using? Table 1

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

(Row 3) shows the simulant that best-fit Monkey 1's performance while using the short-term strategy of rehearsing and monitoring only the sample's trace. Performance on chosen trials hardly decreases as the delay gets longer. This happens because the animal only accepts the test if the sample's trace is available enough, and so the sample's trace when he accepts is always about the same amount above the foil distribution, keeping him equally far from error.

Table 1 (Row 4) shows the simulant that best fit Monkey 1's performance while using the long-term strategy of monitoring activity at the four relevant memory locations. Performance on chosen trials drops sharply at longer delays. This happens because this simulant accepts the memory test if any trace is available enough. With longer delays, the active-enough trace is more often a falsely active foil, and on these occasions the simulant will choose to complete the memory test but will err. One sees from Rows 3 and 4 that formal models can differentiate task strategies and thus can be used to suggest additional research and advance theory.

Monkey 1's performance (Row 1) seems to combine these two patterns. Out to Difficulty Level 3 (a 50sec forgetting interval), his performance on chosen trials stays high, recalling the short-term strategy shown in Row 3. At 100sec, his performance on chosen trials has fallen sharply, recalling the longer-term strategy shown in Row 4. It is possible that this interpretation – of a transition between memory strategies at long delays – could be supported using what is known about the temporal limits of monkeys' working memory (Fobes & King 1982). We offer it only to show that models may illuminate subtle but theoretically important processing differences in animals' performances in uncertainty-monitoring tasks. Actively asking whether the target trace is still available enough, and simply waiting to query the small set of relevant memory traces, are interesting but different capacities.

13.5. Formal modeling and alternative theoretical perspectives

We chose the detection framework as a constructive way to understand animals' performances. However, by assuming trace strengths that quantitatively fade with time, or that wax and wane with serial position, we hope we do not foreclose interest in other theoretical approaches. For example, psychologists have debated whether human recognition memory or memory retrieval can be explained using quantitative continua like trace strength, or whether one must also assume the contextually bound states that are typical of conscious episodic memories and that are often qualitatively present or absent. One could model Hampton's data from Monkey 1 using a more episodic approach, by assuming that at the end of the forgetting interval the animal queried whether it still had qualitatively available the memory of seeing the sample. This approach would also explain in an interesting way why this animal never accepted any trials when, in another condition, Hampton gave him memory tests without having shown him any sample. We point out that Hampton was not willing to endorse this strong a metacognitive interpretation of the performance of Monkey 1. We also point out that this interpretation works less well for the SPR data reported by Smith et al. Their animals might have responded There when they found a contextually bound memory for the previous presentation of the probe item, and might have responded Not There when they did not. However, one still must explain the animals' Uncertain responses which suggest that some quantitative assessment of memory was at work, too. One could speculate that the animals were uncertain whether they had recovered a contextually bound memory, but, as in Hampton's case, the data do not require this interpretation.

In this emerging field, different formal perspectives may each make constructive contributions. Our quantitative approach raises one set of questions about animals' decision-making and criterion-setting processes and about the level these processes have in the psychological system. The qualitative approach would raise another set of questions about animals' episodic memories. In fact, the problem of showing episodic and possibly conscious recollection in animals is an active one in comparative psychology (Menzel 1999; Schwartz et al. 2002).

14. The psychology of uncertain responses and uncertainty-monitoring performances

The facts regarding animals' performances are clear and the SDT descriptions neutral because they do not assume processes (high- or low-level) that would suit one branch of comparative theory but not another. Now though, we come to the difficult point of considering the appropriate psychological description of performance. What is the psychological character of the representations that underlie the SDT continua? How do animals place decision criteria along a continuum and use them? These questions go into areas where different theoretical perspectives are preferred and where theoretical tensions may spark. Yet finding the correct psychological description is one crux of the matter. Therefore, in sections 14.1–14.4, we offer some considerations that help us think psychologically about animals' performances and about their use of Uncertain and Decline responses.

14.1. Animals' Uncertain and Decline responses are not associatively based and are not under stimulus control

One consideration is that a variety of stimulus-based, low-level interpretations of animals' performances are untenable. For example, in the memory-monitoring tasks of Smith et al. (1998), all stimuli, across trials, became targets and foils and were rewarded and non-rewarded following both primary responses. No stimulus cue indicated any response. Only presence or absence in the memory list was relevant. The psychological action in this experiment likely occurred along an internal continuum of subjective trace strength or availability, with animals declining memory tests when probes encountered memories of indeterminate strength. These trial-by-trial assessments of memory strength are profoundly different from the signals available in traditional operant situations, leaving the monkeys' behavior in this task far from traditional senses of stimulus control.

This is already an important interpretation and constraint on psychological theorizing. In essence, the animals are monitoring the contents of memory on each trial. In essence, they are being metacognitive, though one may not wish to load the animals' performances with all the theoretical baggage that this term has with humans.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

A similar conclusion applies to Shields et al.'s (1997) SD task. Behavior in this task also cannot be controlled by absolute stimuli, whether through the generalization gradients surrounding them or the reinforcement histories associated with them. Indeed, the abstractness of the same-different judgment is why many species cannot make it. Here, too, the cues the animals used to decline trials must have been cognitively derived, representing a decision about the (indeterminate) status of the relation assessed on each trial.

14.2. Metacognitive performances and the parsimony embodied in Morgan's Canon

But does this conclusion extend to the perceptual-threshold tasks that began the exploration of animals' metacognitive capacities (sects. 6 and 7 above, and Smith et al. 1995; 1997)? Here one could explain performance using stimulus control or reinforcement history. For example, one might say that stimuli of intermediate density were mildly aversive for being associated with errors and time-outs and that Uncertain responses were conditioned in these stimulus contexts. This low-level explanation has a distinguished pedigree. It defends the principle of parsimony embodied in Morgan's Canon (Morgan 1906, p. 53) that grants animals only simple cognitive capacities. (Remember that Morgan's idea was that one should always interpret an organism's behavior at the lowest possible psychological level.) Thus, given metacognitive-like performances by animals, there is a 100-year-old urge to knock them down, to dismiss them as low-level associative phenomena. Readers may have felt this urge as they read this article. We have several cautions about this theoretical position.

For one thing, in this case the parsimony embodied in Morgan's Canon may be false. The problem is that one cannot interpret the animal's "metacognitive" performance in a vacuum. Humans perform the same way – indeed, the graphs shown in Figure 3 and Figure 4 present some of the strongest existing parallels between human and animal performance. Humans report that they are consciously uncertain and reflexively self-aware as they produce these graphs. And, humans and monkeys share evolutionary pasts, adaptive pressures, and homologous brain structures. Thus, it is unparsimonious to interpret the same graph produced by humans and monkeys in qualitatively different ways – consciously metacognitive versus low-level associative. It uses two opposed behavioral systems to produce the same phenomenon when one might do. Moreover, this duality of interpretation may even be an inappropriate scientific stance. In any other domain, if you showed identical graphs, and then nonetheless offered qualitatively different high-level and low-level interpretations, you would be howled out of the journal. If it were older and younger children, or young and aged adults, or individuals without and with depression, you would have no warrant to do so. Likewise, in the case of humans and animals, you may have no warrant to do so, either.

Given identical performances by humans and animals, one could take the consistent but radical step of claiming that animals and humans are controlled by stimuli and reinforcement in these perceptual tasks, and that humans' introspections and reports of reflexive consciousness and metacognition are a nonfunctional epiphenomenon. This step, though, would deny 100 years of introspections by humans in these tasks, deny the careful understanding that early psychophysicists reached about the Uncertain response in these tasks, and deny the primary source of evidence (self-reports and introspections) that we have for human metacognition and even consciousness. This is a lot to pay for reserving the right to dismiss animal minds.

Fortunately, the issue need not be to either elevate animals or denigrate humans. Rather, given true comparative data and identical performances by humans and animals, our point is just that a reasoned, middle descriptive ground is preferable to the clash between divergent explanatory frameworks at different psychological levels. It is an important principle of comparative research that the kind of integrative parsimony and simplicity of explanation one seeks when explaining the performance of several species in several tasks will be different from the parsimony one seeks when explaining the performance of a single species in a single task. The single-species character of much of comparative psychology has encouraged a sharp focus on low-level, associative kinds of parsimony – the parsimony of Morgan's Canon. Multiple-species studies might foster interest in more integrative kinds of parsimony that could make complementary contributions to theory in the comparative literature. In the present case, a reasoned middle ground could consider both common processing principles across species and acknowledge possible experiential differences across species in whatever way the whole empirical picture warranted.

For another thing, remember that the perceptual-threshold results do not exist alone. Once one knows that animals are using the Uncertain response adaptively to decline derived cognitive states (e.g., indeterminate Same-Different relations, indeterminately available memory traces), another issue of parsimony arises. For now if one explains animals' perceptual performances using a low-level mechanism, but one must explain their memory or same-different performances as a more sophisticated kind of cognitive monitoring, then one grants them two different indeterminacy-resolution systems, one of which is already fairly high-level. In this case one could explain the data more simply just by invoking one indeterminacy-resolution system that applies to indeterminate memory traces, threshold perceptual impressions, ambiguous relational judgments, and possibly many real-world situations, too. Why is such an indeterminacy-resolution system so implausible a thing to assume that animal minds might have benefited from having (sect. 15)? Why is such a system assumed to be more complicated psychologically than if animals learned many sequential reinforcement histories along a perceptual continuum? The history of the comparative literature has led it to answer these questions in one way when analyses of the psychological structure of the different capacities might answer these questions in another way.

Finally, remember also our failure to teach rats to decline trials near their auditory threshold. Rats, pigeons, and other associative species could certainly learn a reinforcement contingency attached to a middle stimulus class between two others and could have a response brought under the control of that class of stimuli. If this were all there is to the threshold task, rats would escape from threshold trials naturally and easily. They do not. This also rules out that these tasks are simply about middle-stimulus avoidance. The

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

threshold tasks seem to be psychologically structured in some way that leaves rats out (insofar as the methods were sufficient to elicit the crucial capacities from them) but leaves humans, monkeys, and dolphins in. Apparently these latter three species are using a capacity at some higher level in the cognitive system that rats access with difficulty.

14.3. Tests of uncertainty monitoring provide inconsistent mappings between stimulus inputs and behavioral outputs, and therefore encourage controlled cognitive processes

Fortunately, one can analyze the psychological structure of the tasks considered in this article, and advance toward a descriptive middle ground that explains indeterminacy-resolution systems across tasks and across species. Shiffrin and Schneider (1977, pp. 167–68) considered the information-processing consequences of the ambiguity that arises from indeterminate cognitive processing. In their description, indeterminate mental representations of stimuli necessarily mapped inconsistently and unreliably onto behavioral responses, making those representations poor indicators of what the organism should do. As a result, higher levels of controlled processing were invoked to adjudicate the indeterminacy. All the uncertainty-monitoring tasks are inconsistently mapped and benefit from controlled processing in this sense.

For example, the perceptual-threshold tasks deliberately challenge the observer's discrimination ability. Given perceptual error, true dense (e.g.) trials and threshold sparse trials will often produce the same subjective impression. Consequently, the impression of density itself will not indicate reliably the appropriate response, and higher-level mediation will be valuable to compare the impression to existing decision criteria to choose a behavioral option. From this perspective, the Uncertain response would be one manifestation of near-threshold resolution processes. It might represent a decision that the trial should be declined because the primary discriminatory process had failed. Note that this description applies to both human and animal cognition. Psychophysical procedures ensure indeterminate stimulus-response mappings and encourage controlled decision-making processes, no matter what the participant species. This realization provides a theoretically gentle way to grant animals' Uncertain responses some of the cognitive sophistication that is due them.

As another example, given memory error in the SPR task, the same trace strength or availability monitored on a trial could be caused by either a There probe or a recently seen Not There probe, and so the trace strength monitored cannot reliably tell the animal what to do. Once again these inconsistent representation-response mappings create the need for controlled resolution processes (see also Gilden & Wilson 1995). In fact, regarding recognition memory specifically, Atkinson and Juola (1974) suggested that the range of indeterminate trace strengths may require qualitatively different, secondary information-processing strategies (i.e., an extended memory search). These indeterminate trace strengths are just the ones that monkeys and humans decline selectively. Their Uncertain responses probably represent a controlled decision, on the threshold of recognition, not to complete the memory test.

To clarify the distinction between decisionally controlled processes and stimulus-controlled processes, consider the conditional discrimination that many humans perform daily – green=go; red=no go. These distinct stimulus input classes eliminate mistakes in perception – stimulus impressions map consistently and reliably onto appropriate responses. Fine discriminations, decision making, and decision criteria are irrelevant. Stimulus and response may associate so strongly that responses are triggered automatically, reflexively, stereotypically, fast, and effortlessly. In fact, this is the point of the worldwide consistent mapping that stoplights offer.

In contrast, imagine that traffic lights gradually morphed from red to green, and that drivers decided whether their light was green enough to go. This situation would be about decision making and decision criteria. It would be about controlled cognitive processing that would be attentional, capacity intensive, and slow. It would also be a nightmare, as perceptual error and self-serving criteria made indeterminate lights seem green enough to go for orthogonal travelers. This kind of decisional task is the one that humans and animals face in the paradigms described in this article. They must decide whether the box is dense or sparse enough to try, the pitch high or low enough, the stimulus relation same or different enough, the memory trace familiar or unavailable enough. Even associative theorists, working in the behaviorist tradition, have given threshold situations like these special attention, for they find that the rules of stimulus control can change there, that animals become minimally informed observers there, and that animals have special problems finding adaptive solutions there (Boneau & Cole 1967; Commons et al. 1991; Davison et al. 1985; Miller et al. 1980; Terman & Terman 1972). In granting this special treatment they follow the classical psychophysicists, who saw the threshold state, and the Uncertain response, as distinctive and complex psychologically.

14.4. Animals may share with humans a theoretically important construal of the threshold and memory-monitoring tasks

Humans make a particular construal of these tasks (i.e., the Dense-Sparse task, the Same-Different task, the serial-probe recognition task) that explains their performance and verbalizations. They accept that the tasks have two primary input classes (dense-sparse, same-different, there-not there). They accept that every trial presents one of these input classes and has a correct answer if they could just discern it. Thus, humans map the two stimulus input classes to the two primary responses, use these responses when they think they know, and reserve the Uncertain response for situations of indeterminacy.

Given this mapping, the Uncertain response alone has no input class associated with it. It is about the status of the primary discriminatory process and about its probable failure. It stands structurally outside the primary discrimination and intrinsically meta to it. For humans, it even attaches to declaratively conscious uncertainty states and, moreover, uncertainty states that are reflexively self-aware in the sense that humans say "I am uncertain." This task construal explains why the Uncertain response feels less stimulus-based, why it can feel like cheating or like mental shirking. It explains why the Uncertain response alone can be omitted from the task's grammar, and why some humans do so by an act of will or bravado. The other two responses cannot be omitted and no human would ever do so. This task

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

construal also confirms the special psychological status that psychophysicists always granted the Uncertain response.

Now, on turning to the animals' performances, leave aside verbalizations, bravado, and consciousness. Animals could still share with humans their abstract construal of the perceptual and memory tasks. In fact, we believe animals do construe these tasks as having two primary input classes and two primary responses that directly map to one another and that exhaust the trial environment, leaving the Uncertain response with no input class associated with it and with a distinctive role in the grammar of the task. The animals make this task construal because we train them, daily remind them, and in some conditions force them to do so.

First, we always train animals extensively in the primary discrimination before giving them an Uncertain response in the task. That response arrives as an optional, extra response, with the two primary input classes and the two primary response classes already established, mapped to one another, and sufficient for performance in the task. Second, we generally provide animals a warmup every day during which they receive easy discrimination trials that ramp up in difficulty toward threshold. During this warmup, it is clear that there are only two stimulus input classes and two useful kinds of responses, and animals only make these two responses. Third, we sometimes run animals in conditions in which they must perform the mature discrimination task without the Uncertain response available. This reinforces again that the task's two input classes and two responses are sufficient for performance in the task. Fourth, animals are also encouraged toward their construal of the task, and toward granting the Uncertain response a special role in the task, by the fact that the Uncertain response alone never earns a direct reward, never earns a timeout penalty, and always has the same neutral function and result in every trial context. Thus, we believe our animals are massively trained in making just the construal of these tasks that we have been describing.

In summary, then, these considerations lay the groundwork for a psychological understanding of humans', monkeys', and a dolphin's successful performances in uncertainty-monitoring tasks. They offer a balanced psychological interpretation of Uncertain responses, granting them the cognitive sophistication they deserve without burdening them with consciousness or with equally heavy behaviorist assumptions.

15. Declarative consciousness

Notice, though, that these considerations do not imply that animals feel uncertainty in these tasks or evaluate within explicit consciousness the status of perception or memory. Regarding the role of declarative consciousness in these tasks, authors (especially ourselves) have been notably cagey. For example, Smith et al. (1998) suggested that

> one could scale back the claims of consciousness while preserving something of the sophisticated, memory-based, flexible, controlled mediational processes that do represent a higher level of choice and decision making and that are needed to explain how monkeys cope with (and escape) indeterminate memory events. (p. 245)

Hampton's (2001) assessment of his analogous memory findings was similarly cautious. On the one hand, he concluded that

> the ability of these monkeys to appropriately decline memory tests when they were unlikely to choose the correct image at test indicates that they know when they remember, a capacity associated with conscious cognition in humans. (p. 5361)

However, Hampton also pointed out (p. 5359), that it is "probably impossible to document subjective, conscious properties of memory in nonverbal animals." Thus, the tack he took was to stress that monkeys have "an important functional parallel with human conscious memory," or an "important functional property of human conscious memory" (p. 5359).

This issue is actually more complex and interesting than it would be if one just prudently denied consciousness to animals as they perform the tasks described here. In fact, our research and that of Hampton raises the cherished idea in cognitive science that cognitive indeterminacy and difficulty inherently elicit higher-level and even conscious modes of cognition and decision making in the organism (Dewey 1934/1980; Gray 1995; James 1890/1952; Karoly 1993; Nelson 1996). James (1890/1952, p. 93) noted that consciousness provides extraneous help to cognition when nerve processes are hesitant. "In rapid, automatic, habitual action it sinks to a minimum." Dewey (1934/1980, p. 59) also argued that in habitual, well-learned behaviors the behavioral impulses are too "smooth and oiled to permit of consciousness of them." Tolman (1932/1967, p. 217) noted that "conscious awareness and ideation tend to arise primarily at moments of conflicting sign-gestalts, conflicting practical differentiations and predictions," such as when the animal is poised on the threshold of a difficult discrimination. We have already encountered Tolman's interesting claim that animals' uncertainty behaviors could operationalize consciousness. Karoly (1993, p. 25) emphasized that uncertain, conflicted conditions are the ones that initiate self-regulation. Gray (1995) described the special neural circuits that may arrest behavior, increase arousal and redirect attention and mental resources toward the causes of difficulty (see also Smith 1995). In the psychophysical literature on humans, too, there is a common view that criterion-setting mechanisms are consciously meta to the primary discriminatory process (Swets et al. 1961; Treisman & Faulkner 1984). Indeed, some psychophysical methods (e.g., the construction of receiver operating characteristic [ROC] curves) depend on humans' ability to obey explicit instructions and consciously set confidence criteria at appropriate levels. Humans comply with all this and often report that their setting of criteria along a continuous dimension is a strategic cognitive process aided by conscious self-regulation.

Now one could still interpret the human and animal results dualistically in this respect, by granting humans declarative, subjective consciousness in these tasks but animals only the unconscious, functional parallels to human conscious cognition. However, we suggest that the claim that difficulty and uncertainty elicit conscious modes of self-regulation is a plausible, principled claim about cognitive architectures generally, one that applies to the human species and to some animal species as well. In fact, this claim has behind it a strong adaptive pressure that might have led as follows to the evolution of working consciousness.

Normally, the systems of human cities (water, food, travel, safety) operate autonomously, reflexively, and automatically in highly trained ways. But given a crisis – a flood,

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

a hurricance, and so forth – the response of the body politic is predictable. A command center is set up. The command center acknowledges that the city's normal, conditioned, reactive mechanisms are not now sufficient. The situation is unfamiliar and could be dangerous. Information must be assimilated and integrated from many sources to adjudicate difficult choices and mediate conflicting goals.

Normally, the behavioral systems of animals (water, food, travel, safety) can also operate autonomously and in their highly trained ways. But there are also times of difficulty – the water-hole dry, a predator interposed between self and young, a canopy-trail home wind damaged, a position of dominance suddenly challenged. These situations cannot be left to habit, to automaticity, or to autonomous gradients of approach and avoidance. The habits do not exist. The situation is novel and untrained. The problem may be multidimensional with difficult choices and conflicting goals. We believe that animals too would have benefited from being able to create the mind's command center for times of uncertainty and difficulty. Working consciousness is ideal for integrating multiple streams of information, for resolving conflicting goals, for coping with the novel and unfamiliar, and for maneuvering on complex optimality surfaces. Working consciousness is the perfect referee for life's close calls. Something like a working consciousness, some cognitive command center, may thus have substantial phylogenetic breadth.

Therefore, we invite colleagues to take seriously the claims of a hundred years of cognitive scientists who noted that the highest levels of information processing, and particularly consciousness, present themselves when difficulty, complexity, and indeterminancy are encountered. Only, we suggest that this is not just a human phenomenon but rather a general functional property of adaptive minds. If you watch an aging cat consider a doubtful leap onto the dryer, you will suspect that what James (1890/1952, p. 93) said is true, "Where indecision is great, as before a dangerous leap, consciousness is agonizingly intense." All the tasks considered in this article place the human and animal participant on just the same doubtful knife-edge of decision, though in the perceptual or cognitive domain. This makes us think that these tasks are well structured and well positioned to elicit these higher-level and possibly conscious regulatory processes in animals. This, of course, is not the same thing as knowing that they do so. It remains an important goal to ask whether the present paradigms can be extended in ways that allow the stronger inference that animals are showing not just the functional features of or parallels to human conscious cognition, but its actual states and feelings.

Appendix 1: Details of simulations and formal models

The same-different task of Shields et al. (1997). Figure 6B showed the results when animals were presented with 13 trial types. Remember that each trial presented the animal with two boxes. If both had Density Level 13, the disparity between them was 0.0 and the trial was Same. As the two boxes had Level 12 and 13 density, or Level 11 and 13 density, and so forth, the trial was an easier and easier Different. One step in modeling was to psychologically scale the objective density-disparity continuum into the subjective-impression continuum that described best animals' perceptual sensitivities in the task. This scaling was done as follows. First, we took the natural logarithm of the 13 relevant pixel densities so that the continuum would obey Weber's law (with the same proportional change in density creating the same psychological change in density). Second, we found the positively signed difference in logarithmic density between each level of disparity and Level 13 (Same). This subtraction translated the scale so that Same trials had a value of 0.0 on the disparity continuum (Level 13 − Level 13), whereas progressively easier Different trials (Level 12 and 13, Level 11 and 13, etc.) had larger and larger positive values. Third, these logarithmic difference values were rescaled through being multiplied by a free parameter (Stretch) in the model to the point that the separations of the 13 values on the continuum correctly reflected animals' sensitivities (assuming, as SDT does, that perceptual error causes subjective impressions of disparity to scatter in unit-normal distributions [SD=1.0] around the objective or expected value of the stimulus event). The stretched scale ran from 0.0 out to 5.913 for the largest disparity presented.

The second step in modeling was to place the Different-Uncertain and Uncertain-Same criteria along this disparity continuum in the way that would most closely mirror the animals' decisional strategy. We searched for these best-fitting placements as follows. We sampled the Different-Uncertain criterion placement at every 1% of the range of the underlying disparity continuum (from 0.0 out to 5.913). We sampled the Uncertain-Same criterion placement at every 1% of the range of the underlying continuum from 0.0 up to the level of the Different-Uncertain criterion placement then in force. That these two criterion points would reverse positions along the disparity continuum makes no sense.

Combining these two steps of modeling, the complete search for the best-fitting parameter configuration of the SDT model evaluated 101 Stretch values, 101 Different-Uncertain criterion placements, and varying numbers of Uncertain-Same criterion placements for a total of 520,251 simulants who each completed 8,000 trials in a virtual version of the monkeys' same-different discrimination. On each trial, the simulant received a trial at 1 of 13 disparity levels, misperceived the disparity according to the perceptual error assumed in SDT, and responded according to its two criterion placements. Summarizing the 8,000 trials, we virtually drew the graph of the simulant's performance profile and compared it mathematically to the performance profile shown in Figure 6B. The criterion of best fit was the sum of the squared deviations (SSD) between the 39 observed and simulated response percentages. The value of SSD for the best-fitting simulant was 810 (39 deviations of around 4% – squared then summed – produced this total). We also calculated the value of a more intuitive fit index, the average absolute deviation (AAD), which is the average amount the simulated response percentages deviate from those observed. The best-fitting AAD was 3.8%. The best-fitting parameter values were 7.07 (Stretch), 0.710 (Same-Uncertain criterion), and 1.360 (Uncertain-Different criterion).

Optimality in the same-different task of Shields et al. (1997). We retained the scaling of the logarithmic disparity axis that fit best animals' sensitivities. These sensitivities define the animals' perceptual limits that cannot be increased in the service of increasing rewards. Along the disparity axis, we surveyed the reward efficiency of strategies that placed the center of the Uncertain response region at 101 places at each 1% increment along the range of the disparity continuum. At each center, we examined the reward efficiency of strategies that gradually widened out the Uncertain response region from having 0 width (zero 1% increments to either side of center) up to 50 width (fifty 1% increments to either side of center). These 5,151 simulants each received 8,000 trials in the virtual SD task, once again receiving 1 of 13 disparities, misperceiving that disparity, and responding according to their two criterion points. These simulants were also subject to the trial times and penalty times of the actual task, so that we could estimate the rewards per minute that each decisional strategy would have received in the actual task and assess the optimality of the monkeys' actual decisional strategy.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

The SPR task of Smith et al. (1998). Figure 7A showed the results when animals were presented with 5 trial types – there were probes that were not in the list and probes that repeated List Items 1, 2, 3, or 4. We assumed that probes on Not There trials contacted on average but with memory variability traces of strength 0.0 (SD=1.0; the leftmost normal curve in Figure 10A), whereas probes on There trials contacted on average but with memory variability trace strengths at the memory sensitivity (*d'*) appropriate to the performance that the animal showed at each serial position (SD=1.0; the four rightward normal curves in Fig. 10A; see MacMillan & Creelman 1991, pp. 209–30). Thus probes of the animal's better (more sensitive) serial positions would encounter stronger traces that lie on average farther from zero on the trace-strength continuum. The first step of the modeling – the scaling of the underlying subjective-impression axis – is given by the animal's sensitivities (*d'*s) so there is no need to stretch or scale the continuum.

Given this underlying continuum, for the forced condition that disallowed the Uncertain response, we evaluated the position of one criterion parameter separating the Not There and There response regions because these were the only two responses granted the animal in that condition. For the optional condition that allowed the Uncertain response, we evaluated the position of two criterion parameters establishing the Not There, Uncertain, and There response regions. The fitting procedure evaluated 61 levels of each criterion point for a total of 226,981 simulants who each completed 8,000 trials in a virtual version of the SPR task. On each trial, the simulant received one of 5 trial types (Not There or a probe of one of four serial positions), assessed (with memory variability) the trace strength this probe item contacted, and responded according to its single or twin criterion placements. Once again we summarized over 8,000 trials the simulant's performance profile and compared it mathematically to the animal's observed performance profile. The criterion of best fit was the sum of the squared deviations (SSD) between the 15 observed and simulated response percentages. The value of SSD for the best-fitting simulant was 210 (15 deviations of about 3% – squared then summed – produced this total). The value of the intuitive fit index, the average absolute deviation (AAD), was 3.1%.

Optimality in the SPR task of Smith et al. (1998). We retained the trace-strength continuum that ran from zero up to higher *d'*s for the animal's more sensitive serial positions, assuming that memory sensitivities were also an information-processing limit that could not be increased in the service of greater rewards. Focusing on the condition with the Uncertain response allowed, we surveyed the reward efficiency of strategies that placed the center of the Uncertain response region at 101 places at each 1% increment along the trace-strength continuum, and, given each center, that widened the Uncertain response region out from having 0 width (zero 1% increments to either side of center) up to 50 width (fifty 1% increments to either side of center). These 5,151 simulants each received 8,000 trials in the virtual SPR task, subject to the trial times, penalty times, and reward structure of the actual task, and responding in accordance with the three response regions in effect.

Monkey 1's performance in the DMTS task of Hampton (2001). We assumed that samples left behind memory impressions at the *d'* (SD=1.0) along the trace-strength continuum appropriate for each delay condition. We assumed that the three foils presented in the memory test were associated with average trace strengths of 0.0 (SD=1.0). Note that memory variability on a trial could cause the sample's trace to be less active than a foil's trace, causing errors in the task, and causing more errors in the task for longer delays with their lower *d'*s. On forced trials, no criterion point applies because the animal must complete all memory tests and it is only a matter of whether he is correct. On optional trials, one criterion line separates the Decline and Accept response regions. The placement of this criterion is the only parameter in this SDT model because here too the scaling of the underlying representational axis is fixed by the *d'*s. The fitting procedure for Monkey 1 evaluated 201 levels of the Decline-Accept criterion, for a total of 201 simulants that each completed 8,000 trials in the virtual DMTS task. The values of SSD and AAD for the best-fitting simulant were, respectively, 320 and 3.7%.

Monkey 2's performance in the DMTS task of Hampton (2001). We modeled the performance of Hampton's Monkey 2 using different procedures that focus on the character of performance when the metacognitive data pattern is not clearly seen. In this case, we assumed that long delays became associated with errors and timeouts and with a reduced urge to complete memory tests. By using this temporal cue, the animal could decline memory tests adaptively at delays, as Monkey 2 and Teller's pigeons did, without consulting any memory trace at all. We assumed that the animal began every trial with a level of Trial-Accept Determination that decayed exponentially as the delay interval transpired. We set this Accept Determination at 3.0 and then we searched a free parameter (Decay) to estimate the steepness of the loss of courage (determination) as time went by. We assumed that at the choice point, the animal's remaining Accept Determination was scattered (SD=1.0 across trials) and that he placed a criterion point on the Accept Determination continuum that let him decline the trials with low remaining determination and accept the trials with high remaining determination. The target data were Monkey 2's 12 data points (Table 1, Row 2). The parameter search for this model involved evaluating two parameters – Decay and the placement of the Decline-Accept Criterion.

The fitting procedure evaluated 21 levels of Decay and 201 levels of the Decline-Accept Criterion, for a total of 4,221 simulants that each completed 8,000 trials in the virtual DMTS task. The best-fitting values of the model were 0.994 (per second) for the decay, and a criterion setting of 1.64 along the Accept-Determination continuum. The values of SSD and AAD for the best-fitting simulant were, respectively, 90 and 2.1%. Table 1 (Row 5) shows that this simulant reproduced well the performance of Monkey 2 (Row 2). The closeness of this fit shows that Monkey 2's performance was essentially not metacognitive, though he may have used a tiny amount of memory monitoring. In fact, when we fit this animal's data making metacognitive assumptions instead, we found a best-fit index that was 12 times as high (i.e., 12 times worse) than when we made the non-metacognitive assumption. The two monkeys in Hampton's experiment showed a clear and interesting individual difference, and illustrate two different kinds of performance in the DMTS task. We have already seen the data pattern of Monkey 2 essentially replicated with pigeons (Teller 1989), and would anticipate similar results if this model were applied to those data.

ACKNOWLEDGMENT

The preparation of this article was supported by the National Institute of Child Health and Human Development (HD-38051).

NOTE

Correspondence concerning this article should be addressed to J. David Smith, Department of Psychology, Park Hall, State University of New York at Buffalo, Buffalo, NY, 14260, or to *psysmith@buffalo.edu*.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Open Peer Commentary

Commentary submitted by the qualified professional readership of this journal will be considered for publication in a later issue as Continuing Commentary on this article. Integrative overviews and syntheses are especially encouraged.

Some sceptical thoughts about metacognition

Derek Browne

Philosophy Department, University of Canterbury, Christchurch, New Zealand. **derek.browne@canterbury.ac.nz**
http://www.phil.canterbury.ac.nz/derek_browne/

Abstract: Metacognitive knowledge of one's own cognitive states is not as useful as is often thought. Differences between cognitive states often come down to differences in their intentional contents. For that reason, differences in behaviour are often explained by differences just in contents of first-order states. Uncertainty need not be a metacognitive condition. First-order interpretations of the target experiments are available.

Consider a subject who is trained to select the red object whenever it is given a choice between a red and a non-red object. Is this behaviour an effect of the subject's capacity to discriminate red from other colours, or is it an effect of the subject's capacity to discriminate between its own experience of seeing red and its experience of seeing other colours? If the latter, then the subject is aware of and can compare some of its own cognitive states: It has some metacognitive capacities. But the simpler, lower-order explanation seems preferable. The subject discriminates differences in colours of physical objects and does so by perceiving those colours, but the subject does not discriminate differences between those perceptual states themselves. The subject has first-order intentional states, through which it is aware of properties of physical objects, relations between those properties, and so on. These first-order states are sufficient to explain a variety of behaviour patterns and differences in behaviour. Colour discrimination tasks do not provide sufficient evidence for metacognition, the ability to be aware of some of one's own first-order states, relations between those states, and so on.

The challenge that experimental studies of metacognition face is to induce behaviour that can only be plausibly explained by reference to second-order states. A confounding factor is that differences between cognitive states (of the same psychological type) are differences in their intentional objects ("contents"). So the explanation of differences in behaviour tends to come down to differences in content. It is the difference between red and green that accounts for the difference between the experience of seeing red and the experience of seeing green. An animal that can discriminate red from green can use that discrimination to drive some of its behaviour. But it is not obvious what additional behavioural responses it will acquire if it acquires in addition the ability to discriminate between the different cognitive states that it is in when it discriminates different colours. That is the challenge facing experimental research into metacognition.

Metacognitive states take lower-order cognitive states as their intentional objects. The contents of the lower-order states are embedded in the contents of the higher-order states. If I am in the first-order state of seeing red, and if I am also in the second-order state of being aware of my seeing red, then the content of the lower-order state (i.e., red) is embedded in the content of the higher-order state. So the standing challenge to metacognition research is to find some pattern of behaviour that could be produced in response to a second-order cognitive state but that is unlikely to be produced in response to its companion first-order cognitive state, given the overlapping content of those two states.

The target article describes some fascinating research that does not go down the well-trodden pathways of the usual studies of deception and false belief. How well does it cope with the difficulty I have described? A salient difference is this: whereas the deception research concerns awareness of the mental states of others, the present research is into a subject's higher-order awareness of its own lower-order states. I will argue that such research is especially vulnerable to the difficulties that I discuss.

The target cognitive state is uncertainty. Uncertainty is not intrinsically a metacognitive state. A subject who is uncertain whether two tones are of same or different pitch is uncertain about the relation between the pitch of the two tones. This is a first-order state, and there is no special difficulty in understanding how it might acquire its own distinctive behavioural signature, such as "runnings back and forth." If the experimental setup allows a third response, an "uncertain" response, does this response give evidence that the subject is monitoring its own first-order states? Is the subject uncertain whether the experience of hearing the first tone is the same as or different than the experience of hearing the second tone? Or is the subject just uncertain whether the first tone is the same as or different than the second tone? We need an argument to decide between the two hypotheses.

There are some striking correspondences between the response patterns of human and nonhuman subjects; that is the first premise of a central inference in the target article. The second premise is that human subjects report that their uncertainty responses are cued by "feelings of uncertainty, doubt, and of not knowing." It is parsimonious to suppose that monkeys who produce the same response patterns as humans produce them from roughly the same cognitive causes.

But there is a problem with this argument. First, it seems to be possible to complete the experimental task without going metacognitive. The task is to judge whether two objects are the same or different: That is a first-order task. The subject feels uncertain whether the two objects are the same or different. This is a first-order state. The content of the uncertainty is not something mental but something nonmental: not whether the experience of hearing one tone is the same as or different than the experience of hearing the other tone, but whether one tone is the same as or different from the other tone. If the task could, in principle, be completed without going metacognitive, what reason is there for thinking that metacognition is in fact involved? Here the postexperimental reports of human subjects become crucial. Humans report being aware of their feelings of uncertainty, and that this uncertainty drives their "uncertain" response. This is not very compelling. When you probe a person to find out why they did something, they become reflective: They go metacognitive. When I report that my feeling of uncertainty caused me to choose the "uncertain" response, my report has a second-order status (it is a report on a first-order state). But this need not challenge the first-order status of the cognitive state that caused the initial response. It was being uncertain that caused the response, not my awareness of being uncertain, just as being certain causes responses of the other two kinds. The reports are metacognitive, but they are not clear evidence that the initial tasks were completed metacognitively.

Furthermore, the fact that a response is not under "stimulus control" is not sufficient to establish that it is under metacognitive control. Sounds and lights and other stimuli must be perceived in order for them to control behaviour cognitively. The perception of the stimulus, not the stimulus as such, drives the behaviour. Suppose some pattern of behaviour is not under the control of (perception of) any current stimulus. This is quite often true of desires. Desires for absent objects can drive behaviour. A thirsty animal sets off in search of water. There is nothing metacognitive here. The animal sets off in a particular direction because it remembers water being over there in the past. There is still nothing metacognitive here. Is there something special about the memory-monitoring tasks to show that responses are driven, not by memory states as such but by a higher-order state that is itself di-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

rected to, or responding to, first-order memory episodes? The situation here is parallel to the one described earlier: Whatever third state of the memory system kicks in when neither remembering that *p* nor remembering that not-*p* kick in, that third state, though still only first order, should be available to drive responses. There is no compelling reason to think that the subject first goes metacognitive to notice that no clear first-order memory is available and then uses that second-order state to drive the "uncertain" response.

On linking comparative metacognition and theory of mind

Josep Call

Max Planck Institute for Evolutionary Anthropology, D-04103, Germany.
call@eva.mpg.de

Abstract: Smith et al.'s article provides a convincing argument for devoting increased research attention to comparative metacognition. However, this increased attention should be complemented with establishing links with comparative theory of mind (ToM) research, which are currently missing. I present a task in which pairs of subjects are presented with incomplete information in an object-choice situation that could be used to establish that link.

Compared to theory of mind (ToM) research in nonhuman animals, comparative metacognition (CM) is an underdeveloped area. Whereas researchers have devoted much attention to the study of mental state attribution in others, much less has been done in the area of mental attribution in the self. I see the target article as an important step in the right direction in readdressing this imbalance. In my opinion, however, there is one thing missing: It falls short of linking ToM and CM.

In recent years our research group has investigated different aspects of ToM in apes (see Call & Tomasello, in press a, b, for recent reviews). Focusing on the area of attribution of visual perception in others, we find that chimpanzees follow the gaze of others to distant locations, including around barriers, and use visual gestures preferentially when others can see them, being sensitive to the face orientation, not just the presence or the frontal orientation of their partners. Moreover, they select pieces of food that others cannot see or have not seen in the past. This sizable amount of information about visual perception in others contrasts with the little we know about what individuals know about their own visual perception. That is, do apes know what they themselves have or have not seen?

To answer this question, Call and Carpenter (2001) presented 2-year-old children, orangutans, and chimpanzees with the following situation. We placed food inside one of two hollow tubes perpendicularly oriented toward the subjects. To get the reward, subjects had to touch the baited tube. There were two baiting conditions. In the visible condition, the experimenter placed the food inside the tube in full view of the subject. In the hidden condition, the experimenter baited one of the tubes but prevented the subjects from witnessing the baiting, so that they knew there was food in one of them but did not know which one. The question was whether subjects would preferentially look inside the tubes before choosing in those conditions in which they had not witnessed the baiting.

Note that this "looking test" shares numerous procedural features with the tests presented by Smith and colleagues, with some significant differences. First, like those studies, our experiment is based on creating uncertainty to see if subjects behave adaptively to eliminate it and maximize their rewards. In the experiments reviewed by Smith et al., this means declining tests, whereas in our paradigm it means seeking additional information. Second, the looking response, like the escape response, is completely optional and never produces a reward or a time-out; instead, touching one of the tubes does. Finally, when apes are presented with containers that do not allow them to look inside, they skip the looking response and simply choose one of the containers. So subjects do not make the optional response when it is not appropriate.

Our results showed that subjects looked into the tubes before choosing more often when they had not seen the baiting. In addition, in 20–30% of the trials subjects made a selection immediately after encountering an empty tube, indicating that seeing the food was not necessary to succeed. In other words, subjects were not simply using the strategy of search until finding the food. Recently, we replicated our results with a different group of orangutans and chimpanzees and extended them to gorillas and bonobos (Call, in press). The positive results with the gorillas were particularly interesting because, unlike other apes, they fail tests of mirror-self-recognition. In contrast with the positive results with apes, dogs failed this test despite being able to retrieve food with high accuracy when they saw where the food was located. Thus, the main finding was that apes including gorillas (but not dogs) presented with a situation with incomplete information about the food location seek additional information to make a correct choice.

In future studies we plan to investigate whether longer delays (see Hampton 2001) between baiting and choosing and higher costs associated with failing to locate the reward would lead to increased looking. We predict that longer delays should foster greater forgetting and therefore increased looking frequencies. Similarly, increasing the cost of failing to locate the reward (e.g., no reward and session termination) should also increase subjects' likelihood of looking even when they have seen the location of the reward. We called this the "passport effect" because people usually double-check the location of their passport before taking an international flight.

After outlining some refinements and potential links with other metacognitive tasks, it is time to turn our attention toward connecting the research on mental state attribution in the self and others. Although currently there is evidence that chimpanzees know what others can and cannot see and what they themselves have seen, these two sets of findings come from different experiments. We plan to investigate both aspects simultaneously by using a modified version of the looking test presented to pairs of apes. On completing the looking test outlined above, pairs of apes will be placed facing each other and presented with two conditions. In the knowledge condition, one chimpanzee (the indicator) is allowed to look inside the tubes and indicate the tube she wants to get, whereas the other chimpanzee (the observer) is only allowed to watch the indicator's behavior but is prevented from looking inside the tubes. After the informed chimpanzee has indicated her choice, the tubes are moved across to the observer chimpanzee to let her choose. In the ignorance condition, neither of the chimpanzees is allowed to look inside the tubes, but the indicator is nevertheless allowed to indicate the tube she wants to get, and again the tubes are moved across to the observer.

If the observer follows the indicator's choices when she is allowed to look inside the tubes but not when she is prevented from doing so, this may indicate that the observer understands that the observer has privileged information only when she was able to look inside the tubes. These data can then be compared to the chimpanzees' performance in the original looking test to see if those chimpanzees who look inside the tubes themselves are also those who trust informed chimpanzees but not ignorant ones.

Finally, I would like to mention another issue that illustrates how closely related the fields of CM and ToM are. Smith et al. pointed out that their results suggest that their subjects have a cognitive sophistication in decision making, not just a set of stimulus-response learned contingencies, but not necessarily declarative consciousness. Similarly, debates on the comparative ToM have often taken this bipolar format. Whereas some researchers propose low-level behavioristic explanations for the observed phenomena, others favor high-level metarepresentational explanations such as false belief attribution. Our position, like Smith et

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

al.'s, falls between these two extreme alternatives (Call 2001; Tomasello & Call 1997; see also Suddendorf & Whiten 2001). We think that apes do more than stimulus response associations, but at the same time, we do not think that they necessarily use metarepresentation of the kind that allows children to solve false belief attribution.

To conclude, CM and ToM have received different amounts of research attention. Smith et al.'s article is an important step in bringing attention to the less investigated topic of CM. However, this increased attention to CM should be complemented with establishing links with ToM. It is only by developing these two areas jointly that we can hope to fully understand the social cognitive processes in humans and animals.

ACKNOWLEDGMENT

I thank Malinda Carpenter for her comments on an earlier version of this commentary.

If metacognition exists in other species, how does it *develop?*

Ruth Campos[a] and Annette Karmiloff-Smith[b]

[a]*Basic Psychology Department, Universidad Autónoma de Madrid, Madrid, 28049, Spain;* [b]*Neurocognitive Development Unit, Institute of Child Health, London, WC1N 1EH, United Kingdom.* **ruth.campos@uam.es**
a.karmilof-smith@ich.ucl.ac.uk

http://www.ich.ucl.ac.uk/ich/html/academicunits/neur_dev/neur_dev_unit.html

Abstract: In this commentary, we raise two issues. First, we argue that in any species, the comparative study of metacognitive abilities must be approached from a developmental perspective and not solely from the adult end state. This makes it possible to explore the trajectories by which different species reach their phenotypic outcome and whether different cognitive systems interact over developmental time. Second, using our research comparing different genetic disorders in humans, we challenge the authors' claim that it is unparsimonious to interpret the same performance in humans and animals in qualitatively different ways, because even the same overt behaviour in different groups of humans can be sustained by different underlying cognitive processes.

Smith et al. are to be particularly commended for their synthesis of the research on animal metacognitive abilities and their stimulating attempt to establish a psychological interpretation of animal metacognitive performance via tasks tapping uncertainty monitoring.

We are particularly concerned with the implications that the authors' proposal has for the study of the progressive development of metacognitive abilities in humans as compared to other species. We note that the authors seem to accept uncritically the fact that all the participants in the various studies they reviewed are adults. In fact, this tends to be the general case in comparative psychology, even when another adult species is directly compared with human infants or children, be it in the domain of metacognition or other cognitive abilities. Such an approach is, in our view, regrettable because, to understand the phenotypic outcome of any species (including human), it is vital to trace the developmental trajectory by which the ability is achieved (Karmiloff-Smith 1998; Paterson et al. 1999; Piaget 1952). Only then can one judge if similar overt *behaviour* is underpinned by similar underlying cognitive processes. Taking a developmental perspective also makes it possible to ascertain whether metacognitive capacities interact over developmental time with other cognitive processes and whether this interaction is analogous in humans and other species. In other words, we believe that cross-species comparisons must involve a developmental perspective both theoretically and empirically. This has proven crucial in the study of metacognitive abilities in humans (Karmiloff-Smith 1992; Perner 1991). Metacognition is not an all or none process. It develops very progressively in human children. The question remains whether this progressive development is the case for those species that the authors claim display aspects of metacognition.

Our second point concerns the issue of parsimony. Normally, of course, one would want to consider similar behaviour between different groups (two groups of humans or comparing humans with other species) as indicating similar cognitive processes. However, our work on normal development has demonstrated that at different ages children may display the same overt behaviour that is nonetheless underpinned by very different internal representations (Karmiloff-Smith 1992). And studies of atypical development suggest that, even when a clinical group attains the same behavioural scores as normal controls, the behaviours are often sustained by very different cognitive and brain processes in the two groups (Grice et al. 2001; Karmiloff-Smith 1998). For example, individuals with Williams syndrome (WS), a rare genetic disorder (Donnai & Karmiloff-Smith 2000), display scores in the normal range on some face-processing tasks. However, in-depth studies revealed that whereas normal controls used configural processing, the participants with WS employed componential processing when looking at faces (Deruelle et al. 1999; Karmiloff-Smith 1998). A second example comes from reading. Normal and WS participants, individually matched on reading scores, turned out to display very different strategies when learning to read new words, suggesting that each group had reached their reading scores via different developmental trajectories (Laing et al. 2001). Such data reveal the need for caution when evaluating similar overt behaviour both within and across species. Would we want to equate parrot speech to human language?

In our view, the answers to both the issues we have raised in this commentary should be sought in the detailed study of the developmental trajectories that lead to the metacognitive abilities which any species displays.

Monitoring without metacognition

Peter Carruthers

Department of Philosophy, University of Maryland, College Park, MD 20742.
pcarruth@umd.edu
http://www.philosophy.umd.edu/people/faculty/pcarruthers/

Abstract: Smith et al. present us with a false dichotomy in explaining their uncertainty data: *Either* the animals' responses are "under the associative control of stimulus cues," *or* the animals must be responding "under the metacognitive control of uncertainty cues." There is a third alternative to consider: one that is genuinely cognitive, neither associative nor stimulus driven, but purely first-order in character. On this alternative the metacognitive reports of humans in these situations reflect states that are interpretative rather than causal in character.

Suppose, first, that beliefs come in *degrees.* Just as desires can be more or less strong or powerful, so something can be believed with greater or lesser confidence. This assumption is routinely made by philosophers. (Some people also claim, however, that there is a distinct kind of ungraded "flat-out" belief, which is dependent on language, and which consists in an explicit *commitment* to assent to a sentence and to think and reason as if that sentence were true. These states are variously called "opinions" or "acceptances" and belong within the so-called dual process models of human cognition. See Cohen 1993; Dennett 1979; Evans & Over 1996; Frankish 1998.) So both animals and humans in the described experimental situations will have a particular *degree* of belief that a tone is high, or that a pattern is dense, for example.

Nothing metacognitive is required for degrees of belief and desire. Having a strong desire doesn't mean that the animal believes *of* itself that it has a desire with a certain strength; rather, it has a desire that is apt to beat out other desires in the competition to control behavior, and that is also apt to have further cognitive and physiological effects of a distinctive sort (e.g., increased heart

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

rate). Equally, having a strong belief doesn't mean that the animal believes itself to have a belief with a certain high degree of warrant; rather, the animal has a belief on which it is more likely to act, all else being equal (and especially given equal strengths of competing desires). And degrees of belief might be realized at different levels of activation in the appropriate neural network, for example.

Suppose, second, that in circumstances in which an animal has conflicting desires of equal strength (to get a food pellet now; to avoid *not* getting a food pellet now – i.e., to avoid a timeout), and where it also has conflicting beliefs of equal strength (the pattern is dense; the pattern isn't dense), the animal will (1) be in a state of some anxiety, which will be experienced as aversive, and (2) will be disposed to engage in behaviors of a sort which tend to eradicate such conflicts of belief (e.g., by attending more closely, changing its angle of view by moving its head from side to side, and so on). Call such a state "an uncertainty state."

Again, nothing metacognitive need be involved. A state of uncertainty is a state caused when the animal is equally disposed to act in two contrary ways, and which has a variety of characteristic further effects of cognitive, affective, physiological, and behavioral sorts (e.g., engaging in information-seeking behaviors). And that an animal engages in behaviors designed (by evolution or by learning) to elicit new information doesn't mean that the animal represents itself as lacking a sufficient degree of belief; it just means that these behaviors are ones that have been sculpted to issue in changed degrees of belief (changes that are apt, in turn, to diminish an aversive state of anxiety).

What is an animal to do in a state of conflict when information-seeking behaviors fail to resolve that conflict? Plainly it needs some heuristic to enable it to reach a decision and move on, or (like Buridan's ass) it will remain frozen. The simplest such heuristic is: *When in a state of that sort, choose at random.* It seems that this is the heuristic many animals (and some humans) employ. A more complex heuristic is: *When in a state of that sort, opt for a less-favored third alternative if you can.* This seems to be the heuristic adopted by some monkeys, dolphins, and humans. Why is this heuristic more complex? Because it requires the animal to represent, and to factor into its decision making, not just the two alternatives between which it faces the initial choice, but also a third option. But what would explain the individual differences in the use of this heuristic? Perhaps differing degrees of risk aversion.

Once again, nothing metacognitive is required for these heuristics to operate. Granted, the animal needs to have some way of telling when it is in a state of the required sort (i.e., the uncertainty state that is the trigger for the heuristic to apply, picked out by the antecedent clauses in the conditionals just mentioned). But this doesn't mean that the animal has to conceptualize the state *as* a state of uncertainty, or as a state in which a conflict of belief remains unresolved. Rather, it just has to have some way of reliably picking out a state which *is* a state of uncertainty. And here its continuing state of anxiety and/or its awareness of its own distinctive information-seeking behaviors (e.g., its side-to-side head movements) would be sufficient.

But how is it, then, that humans will describe these situations in metacognitive terms, if nothing metacognitive is involved in the decision-making process? The answer is: because humans are self-interpretative creatures. Humans will correctly conceptualize the state that they are in as one of uncertainty – that is, as one in which they lack sufficient information to adjudicate between two conflicting beliefs. And they will (again correctly) interpret themselves as choosing as they do because of that uncertainty. But from the fact that their explanations of their feelings and behaviors are metacognitive, it doesn't follow that the decision-making process itself was a metacognitive one. And indeed, there is good reason to think that much of the self-interpretative cognition of humans is *merely* interpretative, not reflective of an underlying metacognitive decision-making process. (For recent reviews of the evidence, see Gazzaniga 1998; Wilson 2002.)

What would it take to demonstrate the existence of genuinely metacognitive processes in animals? It seems unlikely that there are any easy answers to this question. Twenty-five years after Premack and Woodruff (1978) first posed the question whether chimpanzees have a Theory of Mind and represent the mental states of other chimps, there still exists no consensus on an answer. Devising nonlinguistic tests of Theory of Mind ability proved much harder than people initially thought that it would be. There is no reason to expect that tests of animals' ability to represent their own states of mind (including their own states of uncertainty, as such) should be any easier to come by.

Pigeon parallels to human metacognition

Edmund Fantino

Department of Psychology, University of California, San Diego, CA 92093-0109. **efantino@ucsd.edu**

Abstract: The target authors make a strong case for parallels between human and nonhuman metacognition. The case may be bolstered by an appeal to the literatures on commitment and self-control and to that on observing behavior.

The target article presents an interesting array of evidence supporting the suggestion that the study of uncertainty monitoring is a study of metacognition, and that humans and other organisms high on the phylogenetic scale demonstrate uncertainty monitoring and metacognition. Similarities between humans and pigeons have been reported in a series of studies on self-awareness, communication, and insight reported by Robert Epstein and his colleagues (summarized in Epstein 1996). Among other research areas that permit a comparative psychology of behavioral phenomena that may have relevance for metacognition, are those of self-control and of observing behavior, each of which has been the subject of *BBS* treatments (e.g., Dinsmoor 1983; Logue 1988; Rachlin 1995). In each of these cases the results suggest little, if any, qualitative difference across the phylogenetic scale.

The bulk of research on self-control has used humans and pigeons as subjects. Both species use commitment procedures to enhance self-control. In what was probably the first study of self-control in a behavioral laboratory Fantino (1966) found that pigeons generally chose a small immediate reward to a delayed larger reward (although they were sensitive to the experimental contingencies in the sense that, the more advantageous it was to delay, the more likely they were to do so). But the overall level of self-control was extremely low. Ainslie (1974) modified the procedure to give pigeons the opportunity to commit themselves to the larger delayed reward. For three of his pigeons, he found dramatically improved performance. Specifically, the same pigeon that invariably displayed impulsiveness by choosing a small immediate reward to a larger delayed reward committed itself to the larger delayed reward when given the opportunity to make a commitment response (that effectively prevented the impulsive option from occurring). Does commitment imply metacognition in the sense the authors use the term? If so, where commitment procedures are concerned, pigeons and humans flock together.

An issue conceptually related to that of uncertainty monitoring is the potential role uncertainty reduction plays as a reinforcer. This role has been explored in scores of studies using the observing-response procedure first developed by Wyckoff (1952).

In a typical observing-response procedure, reinforcement is provided on either of two schedules of reinforcement but the same stimulus is associated with each. The subject has the opportunity to make an observing response, which may convert this "mixed schedule" into a multiple schedule in which a distinctive stimulus is correlated with each schedule of reinforcement. In other words, when the subject emits a successful observing response, a stimulus correlated with one of the two reinforcement

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

schedules is produced. Thus, a successful observing response informs the subject which schedule is in effect. Observing responses have no other consequence; that is, they are not required for (nor do they affect) primary reinforcement (Fantino 1977). A score of studies have shown that many species, from goldfish (Purdy et al. 1994; Purdy & Peel 1988) through human (Case et al. 1985), observe. Here, again, is a class of behavior that is qualitatively similar across species. Moreover, the species tested share another common attribute in studies of observing: Only the positive stimulus maintains observing. When the only consequence of observing is the possible production of the stimulus associated with extinction or the poorer of two schedules of reinforcement, observing is not maintained (e.g., Fantino & Case [1983] with humans, Purdy et al. [1994] with goldfish). In addition, Fantino and Case (1983) have shown that a stimulus uncorrelated with reinforcement will be preferred to an informative stimulus correlated with the poorer outcome. Relative to "bad news," "no news is good news."

If we accept commitment procedures in a self-control paradigm and information seeking (or avoidance) in the observing-response paradigm, as potential examples of metacognition, then we have further evidence for phylogenetic similarities.

ACKNOWLEDGMENT

Preparation of commentary supported by NIMH Grant MH57127.

Varieties of uncertainty monitoring

John H. Flavell

Department of Psychology, Stanford University, Stanford CA 94305-2130.
flavell@psych.stanford.edu

Abstract: Three types or levels of uncertainty monitoring are distinguished: (1) uncertainty responding but no feelings of uncertainty, (2) conscious feelings of uncertainty, (3) conscious feelings of uncertainty plus reflective awareness of what these feelings are and mean. It is hypothesized that only the first and perhaps the second occur in animals and human infants, whereas all three occur in older humans. Two possible lines of future research are also suggested.

This fascinating article raises questions about the nature and possible varieties of metacognition and uncertainty monitoring, and about which creatures might exhibit which varieties. Human infants and animals clearly respond to uncertainty, sometimes adaptively and sometimes not. The authors cite, in addition to their own experimental data, the compelling example of an aging cat considering whether to attempt a leap up onto a dryer (target article, sect. 15, final para.). Animals in the wild also show similar uncertainty behavior, for example, when considering whether the prey they see might be too big to attack or the predator approaching them too small to fear. Likewise, as graphically shown in the famous "visual cliff" experiments pioneered by Gibson and Walk (1960), human babies may hesitate when encountering a transparent Plexiglas walkway extending over a deep drop-off, some then venturing out onto it and others not. Older children and adults also show uncertainty reactions, of course, again sometimes followed by adaptive responses and sometimes not.

What might be happening psychologically in these situations? One possibility is that the psychological uncertainty reaction, whatever it consists of, does not include any conscious feeling of uncertainty; indeed, it may not include a conscious feeling of any kind. A second possibility is that there is a conscious feeling of uncertainty or hesitancy, but the creature does not reflect on it and consequently does not consciously identify it as being that kind of feeling. A third possibility is that the creature both experiences an uncertainty feeling and is also consciously aware that it is an uncertainty feeling; that is, the creature is unsure, feels unsure, and knows that the feeling it is experiencing is that of being unsure. There are undoubtedly subvarieties of these varieties. Examples might be having only a momentary, semiconscious feeling of uncertainty rather than a sustained, fully conscious one, or having only a vague and fleeting awareness that one is uncertain rather than a clear and enduring one.

We do not know which of these three main possibilities and their imaginable subvarieties occur in nonlinguistic creatures such as the authors' animals, other animal species, and human infants. My guess is that the first and possibly the second may occur in these creatures, but not the third. For older humans, on the other hand, it seems certain that all three occur. Data from a developmental study of comprehension monitoring (Flavell et al. 1981) illustrate some of this variety in the human case. In this study, kindergarten and second-grade children attempted to construct block buildings identical to those of a child confederate, based solely on the confederate's tape-recorded building instructions. After building, they were asked if they thought their building looked exactly like the child instructor's or if it looked different, and whether the instructor had done a good or a bad job of telling them how to make it exactly like hers. Some of the instructions were communicatively very inadequate, containing uninterpretable segments, ambiguous references, contradictions, incomplete directions, or directions that could not be carried out. The older children proved likelier than the younger ones to show spontaneous verbal and nonverbal signs of noticing these inadequacies. As examples, they were likelier to pause, look perplexed, replay the tape, and verbally express their puzzlement. Importantly, the two groups also tended to respond differently to those inadequacies they did detect. Specifically, the older ones were more apt to show that they understood the meaning and implications of the inadequacies by indicating during inquiry that their building might therefore not look exactly like the instructor's and that she had indeed done a bad job of telling them how to make it. A second study showed that the kindergartner's incorrect answers to these inquiry questions were not because of an inability to remember, at the time of inquiry, what the instructions had been. Interestingly, the younger children would often show a brief "flicker" of apparent uncertainty while listening and building (e.g., a momentary pause) followed by no subsequent evidence whatever of problem detection. Whether they could be said even to have had an uncertainty feeling – or any conscious feeling – at that moment appeared questionable, let alone whether they had a conscious, explicit awareness that they did not understand something. As Smith et al. note, a number of researchers currently believe that cognitive monitoring is often automatic and unconscious even in adult humans (Reder 1996).

On a different note, the Smith et al. article raises the possibility of two interesting lines of future research. One is to see if their clever procedure could be adapted for use with human infants and young children. Older babies show at least one adaptive response to uncertainty: they monitor their mother's affective reaction to a potentially scary stimulus for help in deciding whether they should approach it – a phenomenon known to developmental psychologists as "social referencing" (Thompson 1998). Social referencing aside, however, we know next to nothing about the metacognitive capabilities of very young humans. A second research line would be to try to find out what, if anything, animals and human young might understand about the uncertainty reactions of others. For example, could they predict how other subjects would respond on Smith et al.'s uncertainty trials? If they could, might that not suggest that they are capable of a higher, more clearly "meta" level of uncertainty monitoring than can be inferred from present evidence? As Smith et al. indicate (sect. 3, para. 2), theory of mind research tends to focus on what subjects know about others' minds rather than about their own. The two lines of research just proposed would serve to link animal metacognition – the subject of Smith et al.'s article – to human infant metacognition and to animal (and human) theory of mind. It is clear from their article that the authors would favor such links.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Dolphins on the witness stand? The comparative psychology of strategic memory regulation

Morris Goldsmith and Asher Koriat

Department of Psychology, University of Haifa, Haifa, 31905 Israel.
mgold@research.haifa.ac.il **akoriat@research.haifa.ac.il**
http://iipdm.haifa.ac.il

Abstract: Smith et al. show that monkeys and dolphins can respond adaptively under conditions of uncertainty, suggesting that they monitor subjective uncertainty and control their behavior accordingly. Drawing on our own work with humans on the strategic regulation of memory reporting, we argue that, so far, the distinction between monitoring and control has not been addressed sufficiently in metacognitive animal research.

In their stimulating target article, Smith et al. argue that humans are not the only animals that metacognize. Seemingly, monkeys and dolphins also possess at least a rudimentary ability to monitor their own state of knowledge and control their behavior accordingly. In this commentary, we highlight the distinction between *monitoring* and *control* (a core distinction in metacognition; Nelson & Narens 1990), drawing on our own work with humans on the strategic regulation of memory accuracy performance (e.g., Goldsmith & Koriat 1999; Goldsmith et al. 2002; Koriat & Goldsmith 1994; 1996b; Koriat et al. 2001).

To draw an analogy between Smith et al.'s work and our own, consider the situation of courtroom testimony. In attempting to "tell the whole truth and nothing but the truth," witnesses must regulate what they report from memory, volunteering information that is likely to be correct and withholding information that is likely to be wrong. In our studies, we found that when allowed to withhold answers about which they feel unsure (*free-report* conditions), people can enhance substantially the accuracy of what they choose to report relative to forced report (Koriat & Goldsmith, 1994; 1996b; Koriat et al. 2001). Moreover, they exercise this option strategically, enhancing accuracy even further (by withholding more answers) when given stronger incentives to do so.

Could a monkey or a dolphin "witness" also regulate its memory reporting effectively? Apparently yes. Treating the "uncertain" response as a withheld ("don't know") response: (1) Monkeys and dolphins apparently can respond "don't know" when they feel uncertain. (2) When they choose to make a primary response (volunteer an answer), they are more likely to be right than when they opt for a "don't know" response. (3) Consequently, they can increase the accuracy of the chosen primary responses (volunteered answers) by selective use of the "don't know" response (see especially Monkey Baker's impressive performance in Smith et al. 1998).

Despite these similarities, however, before we can conclude that dolphins' and monkeys' metacognitive processes parallel those of humans, more refined distinctions need to be made than are afforded by the signal-detection-theory (SDT) approach endorsed in the target article.

In our model (Koriat & Goldsmith 1996b; for an extension, see Goldsmith et al. 2002), we assume that people first retrieve information, then assess the likelihood that the information is correct (monitoring), and then decide whether or not to report it (control). This decision depends not only on the monitoring output (confidence), but also on the perceived incentives for accurate reporting. This model is similar to the one implied in the target article. However, based on this model, we identified four distinct factors, alone or in concert, that contribute to free-report memory performance: (1) *memory retention* – the amount and quality of information accessible in memory, (2) *monitoring effectiveness* – the ability to discriminate between correct and incorrect information that comes to mind, (3) *control policy* (report criterion) – the liberality or strictness of the criterion that is adopted for volunteering or withholding answers, and (4) *control sensitivity* – the extent to which the control decision (volunteer/withhold) is in fact based on the monitoring output. In testing the model (Koriat & Goldsmith 1996b), we developed an experimental paradigm and assessment methodology that allows each of the various cognitive and metacognitive factors to be measured independently (as well as free-report quantity and accuracy performance), by obtaining forced-report and free-report responses to each memory item and a confidence rating for each forced-report answer (see the QAP assessment procedure in Koriat & Goldsmith 1996a; 1996b).

Note that Smith et al.'s claims specifically concern monitoring – the ability to know that one knows. Thus, an important challenge facing students of animal metacognition is to distinguish between monitoring and retention, on the one hand, and between monitoring and control, on the other. Consider first the distinction between monitoring and retention. Clearly, one can have poor retention and yet be able to distinguish between correct and incorrect answers that come to mind. Conversely, one can remember quite a lot, yet be unable to screen out wrong information. In our work (Koriat & Goldsmith 1996b), we showed that output-bound memory accuracy (the proportion of volunteered answers that are correct) depends crucially on monitoring effectiveness, independent of retention.

Turning to the distinction between monitoring and control, people may adopt a strict or liberal report criterion independent of monitoring (confidence level), depending on the relative weight given to the (competing) incentives for complete versus accurate reporting (Koriat & Goldsmith 1994; 1996b). Moreover, it is not a foregone conclusion that the control decision will be based on the monitoring output: Although with undergraduate participants, control sensitivity is virtually at ceiling (Koriat & Goldsmith 1996b), ongoing research with elderly people (Pansky et al. 2002) and clinical populations (Danion et al. 2001; Koren et al. 2001), suggests that this factor too may be nontrivial. In general, then, monitoring (confidence) does not dictate specific report decisions, nor do such decisions exclusively reflect monitoring.

As discussed elsewhere (Koriat & Goldsmith 1996a; 1996b), the foregoing distinctions become blurred when the signal-detection framework is adopted. For example, in the "old/new" memory paradigm to which signal-detection methods are typically applied, "control" is isolated in terms of the parameter beta, but "memory retention" (overall memory strength) and "monitoring effectiveness" (the extent to which the participants' confidence distinguishes "old" from "new" items) cannot be operationally or conceptually separated: Both are equally valid interpretations of d′ (see, e.g., Banks 1970; Lockhart & Murdock 1970). Moreover, control sensitivity is a nonissue: It is axiomatic in SDT that the control decision (e.g., old/new) is based on memory strength, which is generally equated with confidence. This may explain why, notwithstanding the cleverness and elegance of Smith et al.'s methodology, it only taps monitoring *via the control decision*, rather than independently (see also Higham 2002).

To support the independent status of subjective monitoring in animals and refute behavioristic interpretations, it is not sufficient to show that animals behave *as if* they monitor their own knowledge. In fact, in the absence of verbal measures of confidence, it is crucial to demonstrate that they can behave *otherwise*, that is, to demonstrate a dissociation between the monitoring and control functions. Notwithstanding the methodological challenges involved, Smith et al.'s (1995) study, described in the target article, offers room for optimism: If noninstrumental behavioral indices such as hesitancy, slowing, and wavering could be developed into reliable and sensitive measures of animal confidence, it would be possible to examine separately the effectiveness of memory monitoring and of control. For example, to isolate the control policy, one could manipulate the incentive for accurate reporting, and examine whether, as one would expect, the manipulation affects the animal's report criterion without affecting confidence. With regard to control sensitivity, one might ask whether the apparent absence of animal metamemory in Shields's (1999) prospective memory-monitoring task stems from an inability to monitor or, rather, from the animal's failure to control its responding on the basis of

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

its monitoring. More generally, one could investigate the conditions under which animals (including humans) override their monitoring, responding on other bases (e.g., desperation, compulsion, perseverance, drive). We believe that the dissociation between monitoring and control may offer another small window through which to examine issues concerning the role of conscious awareness in cognition, metacognition, and behavior.

ACKNOWLEDGMENT

We gratefully acknowledge support of this project by the German Federal Ministry of Education and Research (BMBF) within the framework of German-Israeli Project Cooperation (DIP).

Significant uncertainty is common in nature

Donald R. Griffin

Concord Field Station, Harvard University, Bedford, MA 01730.
griffin@fas.harvard.edu

Abstract: In animals' natural lives, uncertainty is normal; and certainty, exceptional. Evaluating ambiguous information is essential for survival: Does what is seen, heard, or smelled mean danger? Does that gesture mean aggression or fear? Is he confident or uncertain? If they are conscious of anything, the content of animals' awareness probably includes crucial uncertainties, both their own and those of others.

Our scientific thinking about uncertainty monitoring and metacognition in animals tends to be framed in terms of laboratory animals, whose lives have much more certainty and much less stress and danger than is usual under natural conditions. A broader consideration of the behavior and cognition of animals during their normal lives can add significant evidence about awareness of uncertainty. Smith et al. recognize that "Working consciousness is the perfect referee for life's close calls. . . . [and] . . . may thus have substantial phylogenetic breadth" (sect. 15, second last para.). A comparable view has been expressed by Dickinson and Balleine (2000) concerning the goal-directed action by laboratory rats.

The content of animal consciousness is doubtless very simple and limited in comparison with ours; but it can nevertheless be important to the animal for its survival.

Under natural conditions animals must wonder whether a marginally visible movement of vegetation or a barely audible rustling means that a dangerous predator is present. But wind and other harmless events cause very similar rustlings. Seed-eating animals probably wonder whether a particular speck on the ground is a pebble or a partly buried seed. Social signals often have uncertain meaning: Does that slight movement of a furry shoulder mean he will attack, or does that brief glance mean that she will be friendly? Which of those infant distress calls is from my baby? Such uncertainties are often vitally important for the animal and are therefore likely to be consciously experienced and evaluated.

I suggest that we adopt the potentially testable working hypothesis that many animals are consciously aware of whatever is critically important in their lives, and that simple perceptual consciousness is a core function of central nervous systems. Selecting actions the animal believes will obtain what it wants or avoid what it dislikes or fears is an efficient way to use a central nervous system. This ability is adaptive because it makes appropriate decisions more likely and thus increases the animal's evolutionary fitness. If animals are aware of anything, the many uncertainties that are critical for survival must often require conscious attention.

A major obstacle to evaluating this hypothesis is the widespread opinion that it is impossible to determine with absolute certainty whether an animal is or is not conscious. Yet we seldom if ever demand perfect proof before evaluating imperfect evidence about other difficult questions in the behavioral sciences, so that this double standard is a form of paralytic perfectionism that discourages research. Furthermore, there is now abundant evidence of nonhuman cognition and consciousness, as reviewed by Heyes and Huber (2000), Griffin (2001), and Bekoff et al. (2002). Considerable information about the content of consciousness is readily available. We make inferences about the conscious states of our human companions by interpreting their communicative behavior, both linguistic and nonverbal. This is increasingly feasible with animals as more is learned about the versatility of their communication. They often appear to be communicating their conscious experiences, which amounts to declarative consciousness, even though their communication systems differ from human language.

Smith et al. are concerned primarily with metacognitive thinking about uncertainty itself in contrast to thinking about alternative possibilities on the basis of imperfect information. I suspect that many animals are keenly aware of uncertainties about the meaning of sensory information that may or may not mean danger or opportunity; but whether they think about uncertainty as an abstract concept is much more difficult to judge. Perhaps we should search for communicative behavior that reports awareness of uncertainty itself. This might occur naturally, once we learn where to look for it, or it might be instilled by extensions of the types of experiments reviewed in the target article. In many challenging situations when animals are uncertain what to do, they actively seek better information, peering, listening, tasting, probing, or sniffing. Prey animals often show greater anxiety and caution when moving through thick vegetation where predators are more difficult to see than out in the open. And dangerous predators are sometimes inspected cautiously, apparently in search of an indication of the intention to attack. Sometimes this information quest entails exchanging communicative signals with other animals.

One example is the exchange of symbolic gestures by swarming honeybees recently reviewed by Griffin (2001). When it is necessary that the swarm find a suitable cavity into which they can move, some scout bees that have located reasonably suitable cavities report their distance, direction, and desirability by the same symbolic gestures ordinarily used to report the location of food sources. Some follow dances of a sister that describe a better cavity. Occasionally, the first bee then changes her dance message to that describing the better cavity about which she has learned as a dance follower. Sometimes this occurs *without first inspecting the second cavity herself.* This appears to be an example of the "substantial phylogenetic breadth" of working consciousness, recognized by Smith et al., applied in a situation where the bees are uncertain which is the better cavity and are seeking additional information before making this vitally important decision.

Metacognition as evidence for explicit representation in nonhumans

Robert Russell Hampton

Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda, MD 20892. **robert@ln.nimh.nih.gov**

Abstract: Metacognition is either *direct,* as when information is recalled before making a confidence judgment, or *indirect,* as when the probability of successful future retrieval is determined inferentially. *Direct* metacognition may require an explicit mental representation as its object and can only be demonstrated under specific experimental circumstances. Other forms of metacognition can be based on publicly observable stimuli rather than introspection.

Metacognition requires two distinct components, an object-level mental process, such as a memory, and a meta-level, or executive, process that monitors the object-level process (Nelson & Narens 1996). In some cases, the meta process has direct access to an explicit representation at the object level (e.g., Koriat 1996), but in many other cases monitoring is indirect or inferential (Flavell 1979). Contrast the following two situations requiring a metacog-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

nitive judgment: (1) A colleague asks if you remember how much money President Bush suggested allocating to hydrogen car research in his state of the union speech, and (2) a friend calls and asks whether you can answer a question his eight-year-old child has about psychology. In the first case, one would surely check the contents of memory and determine whether one retrieves a memory of this piece of Bush's speech. The meta-level judgment of knowing thus depends on the success or failure of attempted access to the primary explicit or declarative memory (hereafter *direct* metacognition). By contrast, in the second case your friend has not asked you to retrieve a specific memory. But if you are an expert in psychology you might feel confident that you can answer the question of an eight-year-old. Your confidence is not dependent on a direct evaluation of the contents of your memory, but rather on your history of expertise, past ability to answer such questions, and assessment of the intellectual capacity of an eight-year-old (hereafter *indirect* metacognition). It is significant that in this second case your friend could make a judgment about your ability to answer correctly that is about as accurate as your own judgment. This would not be true if you were directly accessing a specific explicit memory, in which case you as the introspecting individual would have a distinct advantage in estimating the probability of a correct response.

Evidence for *direct* metacognition in nonhumans supports the position that nonhumans have declarative or explicit mental representations (Hampton 2001). It is possible to demonstrate direct metacognition only under circumstances that rule out alternative, potentially subtle, sources of information that could inform the meta-level process. For the purpose of testing for explicit representation in nonhumans, the strategic decision to avoid some tests (a behavioral choice driven by the meta-level process) is not the phenomenon of primary interest; it is merely the behavioral output that demonstrates explicit mental representation. To demonstrate that an object-level representation is explicit, and therefore capable of supporting direct metacognition, most of the experimental effort is spent eliminating indirect sources of information for the meta process. In the absence of indirect sources of information, we can conclude that the meta process is monitoring the object-level representation directly. Thus, direct metacognition may tell us more about the nature of the object-level representation than it does about the meta-level processes. In investigations of the meta-level process per se, the painstaking avoidance of subtle cueing may be less relevant, and an indirect metacognition paradigm suitable. Indeed, in educational settings students are encouraged to examine manifest features of the study material, such as quantity, complexity, and format, then to allocate study time appropriately (a meta-level process) on the basis of these publicly observable properties of the to-be-learned material. Here, the strategies used are of primary interest, whereas the basis on which particular strategies are chosen is less critical. Experiments examining whether or not nonhumans can alter the character of information processing strategically might well emulate this emphasis on overt cues informing the meta-level process.

In making sense of the work reviewed in the target article, the distinction between *direct* and *indirect* metacognition is critical. The evidence reviewed clearly indicates that some nonhumans are capable of some form of metacognition, shown by the fact that monkeys, dolphins, and sometimes pigeons can strategically avoid tests when they are unlikely to perform correctly. However, the excitement with which one responds to these findings may depend on what stimuli control the animals' choice to avoid a test. For example, if the duration of the delay since study controls the choice to take or avoid a memory test, this would be a clear case of *indirect* metacognition. By timing the delay on each trial, the animals indirectly monitor their memory state and avoid tests when memory is likely to be weak. But such a result would not indicate declarative or explicit representation of knowledge in a nonhuman, because the ability depended on an external stimulus rather than on a judgment based on introspection.

With respect to the question of explicit representation, the procedures used by Inman and Shettleworth (1999) and Hampton (2001) are distinct from the other procedures described in the target article. In both studies, the subjects were required to make *prospective* judgments of knowing or certainty, *in the absence of the primary test stimuli.* By ruling out sources of indirect information for the metacognitive judgment, this procedure constitutes a strict test of the hypothesis that the metacognitive judgment is based on introspection directed at explicit mental representations. As one example, under these conditions subjects cannot use their own vacillation or hesitation (Tolman's "runnings back and forth" referred to by the authors) as a basis for the metacognitive response, because they are not engaged in the test at the time they make the metacognitive judgment. Because these ancillary behaviors could be the result *of* the subject's inability to rapidly complete the test, rather than a reflection of the subject's knowledge *that* it (he/she) cannot complete the test, such behaviors can constitute an overt cue that could control the choice to avoid a test. From a comparative perspective, it is striking that under strict test conditions pigeons have so far failed to show direct metacognition and explicit representation, but monkeys have succeeded (Hampton 2001; Inman & Shettleworth 1999; J. E. Sutton & S. J. Shettleworth, personal communication, February 12, 2003). However, under less strict conditions, where the pigeons made the metacognitive judgment in the presence of the test stimuli, they behaved in a way much more consistent with metacognition. Although more work remains to be done, these results point to the possibility that animals as diverse as humans, pigeons, and monkeys are capable of behavior that is superficially similar and apparently metacognitive. However, monkeys and humans, but not pigeons, can accomplish this behavior through introspection directed at explicit mental representations. This does not deny pigeons' minds, but it does suggest that carefully controlled experiments show their minds to differ from ours.

ACKNOWLEDGMENT

Preparation of this commentary was supported by NIMH IRP.

Parsimonious explanations and wider evolutionary consequences

James E. King

Department of Psychology, University of Arizona, Tucson, AZ 85721.
kingj@u.arizona.edu

Abstract: The uncertainty response adds an important new dimension to conventional animal learning and memory studies. Although the uncertainty response by monkeys and dolphins resembled that of humans, parsimony alone does not necessarily indicate that the monkeys and dolphins had a full self-awareness. However, the uncertain response may be an index of an evolutionary precursor to full self-awareness of uncertainty and a theory of mind.

Studies of two-choice problems including memory, discrimination, and matching tasks have a long and venerable history in animal learning research. However, interest in these problems has been recently eclipsed by more complex and exotic variations, particularly in the realm of primate studies where research on complex learning, theory of mind problems, and language have largely replaced the old discrimination and learning-set paradigms (see Tomasello & Call 1997).

Smith et al. have described an interesting and potentially important new dimension to simple learning and memory problems. The uncertain response yields a certain outcome that is less attractive than that following a correct response but more attractive than the outcome of an incorrect response. The most important findings of the research reviewed by Smith et al. was not simply that uncertain responses in rhesus monkeys and dolphins increased when the demands on memory or perceptual sensitivity

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

were at a level that led to errors during conventional testing without the possibility of uncertain responses. Instead, the imaginative use of SDT to monkey performance showed that probability of uncertain responses peaked at parameter values where the independently verified problem difficulty was greatest. Rhesus monkeys are apparently able to make remarkably accurate quantitative assessments of their own uncertainty and express that quantitative information with objective responses.

The value of a new technique or procedure is largely determined by the extent to which it suggests interesting new questions extending beyond the original paradigm and lends itself to experiments or observations that answer those questions. The accurate use of the uncertain response by monkeys and dolphins leads Smith et al. to suggest that it is a manifestation of a simple form of metacognition and possibly self-awareness. The central issue is then whether a simpler explanation would work. Smith et al. dismiss the possibility that the uncertain response could be conditional discrimination based on emotional responses accompanying perceptually difficult discriminations or lack of memory. The lack of a specific external stimulus class associated with the uncertainty emotional response is given as a justification for rejecting the discriminative stimulus interpretation. This rejection may be premature. Abundant evidence exists that internal states can serve as discriminative stimuli. Rats have been regularly shown to be capable of learning overt motor discriminations based on hunger and thirst (e.g., Bolles & Petrinovich 1954) or fear (Rosellini & Terris 1976).

Smith et al. somewhat misconstrue Morgan's *Canon* when they state that it grants animals only simple cognitive capacities. The *Canon* is not an empirical declaration that simple explanations of animal behavior and learning are most likely to be correct. Instead, it states an arbitrary, but entirely modest and reasonable procedural principle, namely, that animal behavior should be interpreted in terms of the simplest explanation consistent with all available evidence. In fact, Morgan himself believed that animal behavior could be interpreted in terms of intentionalist states and that the animal mind could not only be affected by evolution but also be an active agent for evolutionary change (Costell 1993). It is therefore unfortunate that Lloyd Morgan has been commonly associated with the view that the animal mind must be no more than a Cartesian automaton.

The core of Smith et al.'s argument is that a principle of cross-species parsimony supports the attribution of metacognition to rhesus monkeys and dolphins. Their reasoning is that because humans in the uncertainty experiments were "consciously uncertain and reflexively self-aware" (target article, sect. 14.2) when using the uncertainty response, it would then be unparsimonious to interpret a similar use of the uncertainty response by rhesus monkeys and dolphins in terms of a completely different and simpler process. A similar idea about parsimonious explanations for similar behaviors across species was described as "evolutionary parsimony" by de Waal (1996). However, parsimony can be a slippery concept. A lack of parsimony entailed in attributing a behavior in one species to mechanism A and a similar behavior in another species to mechanism B would be fully justified only if A and B were qualitatively separate phenomena. The comparison across species becomes trickier and more interesting when A and B differ only quantitatively, or when A and B share some but not all attributes, or when A is somehow more fundamental and evolutionarily more primitive than B. In the later case, A may be a necessary accompaniment of B. An example of this type of evolutionary hierarchy was illustrated in de Waal's studies of possible evolutionary precursors to the sense of morality and justice in humans. Chimpanzees display delayed reciprocity of friendly as well as unfriendly acts in social situations (de Waal & Luttrell 1988). Furthermore, chimpanzees showed an understanding of the future consequences of their current behaviors (prescriptive behaviors) (de Waal 1991). These behaviors are not instances of a human sense of justice and morality. However, a convincing argument could be made that the capability for social reciprocity and prescriptive behaviors were necessary precursors for the later evolution of behaviors reflecting human senses of justice and morality.

By similar reasoning, an argument could be made that the adept use of uncertainty responses by rhesus monkeys and dolphins indicates an ability that must have evolved before self-awareness and ultimately a theory of mind could have evolved. Smith et al. did suggest this possibility. Uncertainty response problems have a logical relationship to a class of theory of mind experiments in which a subject chooses to receive reward relevant information from one of two human informants. The subject is given information indicating that one informant has the relevant information while the other one does not. Consistent choice to receive information from the knowledgeable informant indicates an ability to monitor knowledge of someone else – a logical and important generalization of the uncertainty response. Although results on this task are not entirely consistent, chimpanzees have on occasion been successful in perceiving the presumed awareness of human models (e.g., Povinelli et al. 1990). However, results with macaque monkeys have been generally negative (Povinelli et al. 1991). This extension from monitoring of self- knowledge to monitoring of the knowledge of others would be an important extension of the research described by Smith et al.

Can we be too uncertain about uncertainty responses?

Lori Marino

Neuroscience and Behavioral Biology Program, and Living Links Center for the Advanced Study of Ape and Human Evolution, Department of Psychology, Emory University, Atlanta, GA 30322. **lmarino@emory.edu**
http://www.emory.edu/LIVING_LINKS/i/people/marino.html

Abstract: Smith et al. propose that the most parsimonious explanation for identical responses of humans and nonhumans under the same conditions is not always the simplest cognitive explanation but could be the one that has the most logical consistency across species. The authors provide convincing evidence and a reasonable argument for declarative consciousness as a shared psychological property in humans, monkeys, and dolphins.

Smith et al. are to be commended for presenting a clear-sighted and logical approach to an area of research – metacognition in other species – that, in less capable hands, would easily slip into a quagmire of inchoate ideas. One of the strengths of their paper is that they present a testable and highly accommodating theoretical framework for research on metacognition in both humans and nonhumans by Nelson and Narens (1990) and then proceed to show how a domain of research on uncertainty monitoring in humans and other animals fits into this scheme. This approach anchors the research they present in a common framework. The authors' quantitative and formal modeling approach provides an extremely useful way to map cross-species abilities onto the same topography and produce testable predictions. Another advantage of the paradigm the authors' promote is that outcomes can remain descriptive; the interpretation of these outcomes is not entirely reliant on adopting one specific theoretical explanation. This leads me to the main focus of my commentary, which is on the authors' discussion of the role of parsimony in explaining the findings they review and the kinds of questions these outcomes lead me to think about regarding how we view the domain of consciousness in humans and other animals.

Given identical findings for humans and nonhumans on metacognitive tasks in a truly comparative setting, Smith et al. argue that the principle of parsimony demands that we consider a "reasoned, middle ground" that does not always invoke the simplest cognitive explanation, but rather the most consistently explanatory one across species. The authors describe this approach as integrative parsimony. Integrative parsimony involves acknowledging the generally heterogeneous nature of psychological mechanisms.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Therefore, when interpreting the same data in different species, we should acknowledge *both* common processing mechanisms (especially when two species are phylogenetically close and share homologous brain structures) on some levels and possible differences on other levels.

This brings us to consciousness and self-awareness. Smith et al. are appropriately cautious about going beyond purely descriptive statements that monkeys and dolphins show functional parallels to declarative human consciousness. They are understandably conservative about making explicit pronouncements of declarative consciousness in other species. On one level, it is a truth that no matter how similar a nonhuman (or even another human) individual's responses are to our own, we can never enter into the subjective realm of that individual and be absolutely sure, on the personal level, that the individual is experiencing the same psychological states as we are. However, the authors also imply that their findings demonstrate functional equivalence of human, monkey, and dolphin responses but not "actual states and feelings." A word of caution comes to mind. By stopping short of full acknowledgment of declarative consciousness in other animals within such a rigorous paradigm, it would be prudent to consider whether it is tenable for responses that *look* metacognitive to work *without* declarative consciousness.

Smith et al. make a very convincing argument that it would be reasonable to expect "something like a working consciousness" to be a highly adaptive general property of many species. Humans and nonhumans engage in subconscious automatic behaviors when situations are predictable and require mainly highly trained responses. However, there are unpredictable situations that require more "on-line" types of information processing, that is, consciousness. Smith et al. give a nice anecdotal example of an elderly cat deciding whether to jump up onto a high platform like a dryer. Now, if the cat were asked to report on whether it was uncertain in those circumstances and did so, is not the cat therefore offering unequivocal evidence of declarative consciousness? The ability to comment on one's mental states is, by definition, metacognitive and declarative. It seems to me that this is equivalent to the situation in which monkeys and dolphins "report" uncertainty. (There is even ancillary evidence from the dolphin that the animal displayed typical approach-avoidance behavior during difficult trials.)

My point here is that, surely, we have to be careful not to make overly strong inferences about the subjective experience of other animals but also be equally cautious about trying to place nonhuman behavior in a theoretical space that may not exist. Could other species display identical response patterns to humans under the same highly rigorous conditions and not be drawing on the same psychological mechanisms? Are we, in some way, asking other animals to be even more metacognitive than we are? To be meta-metacognitive?

The same kinds of questions might be asked about the findings that great apes and dolphins use mirrors to investigate parts of their own body (Parker et al. 1994; Reiss & Marino 2001b). There are clearly psychological differences among humans, great apes, and dolphins. Yet we should not let recognition of those inevitable differences be the impetus for unnecessarily unparsimonious explanations whereby we force ourselves to pose different psychological explanations for identical responses under the same conceptual conditions. In Reiss and Marino (2001b) we conclude that mirror self-recognition in bottlenose dolphins is a case of cognitive convergence with primates, implying that the same or similar adaptive psychological mechanisms can evolve in phylogenetically distant species.

In conclusion, Smith et al. may have accomplished more than they intended in their paper. They provide not only substantial evidence but also a cogent argument for declarative consciousness as a shared psychological property in humans, monkeys, and dolphins. That may indeed be the most parsimonious explanation for the findings they describe.

Animals show monitoring, but does monitoring imply awareness?

Giuliana Mazzoni

Department of Psychology, Seton Hall University, South Orange, NJ 07079.
mazzongi@shu.edu

Abstract: The very clever studies reviewed by Smith et al. convincingly demonstrate metacognitive skills in animals. However, interpreting the findings on metacognitive monitoring as showing conscious cognitive processes in animals is not warranted, because some metacognitive monitoring observed in humans appear to be automatic rather than controlled.

In recent years the idea that cognitive processes play an important role in animal learning and behavior has gained vast consensus (e.g., Lovibond & Shanks 2002; Miller & Oberling 1998), and a growing body of data support the idea that performance on a number of tasks is better explained by higher-order cognitive processes than by simple association (for very recent examples, see Bekoff et al. 2002; Colombo & Frost 2001; Morris 2002; Terrace et al. 2003). Smith et al. extend this by arguing that at least some mammals display metacognitive abilities. They go on to argue that this demonstrates these mammals "have functional features of or parallels to human conscious cognition" (target article, Abstract). In this commentary, I argue that they make a convincing case for the first of these propositions, but not the second.

Studying metacognition in animals might be considered as an insurmountable challenge by many researchers. The procedures described in the target article prove the contrary. They show that simple and very clever experimental designs can provide rather compelling data supporting the hypothesis that rhesus monkeys and at least one dolphin are able to "decide" to skip a test when the test is too challenging for them and the outcome of their performance is not certain. The possibility of avoiding a test that will give uncertain and probably wrong results is a situation similar to the one described by Koriat and Goldsmith (1996) with university students. In a series of studies, it was observed that students selectively avoided answering questions that would lead to wrong responses when the cost of the errors was high. Performance improved significantly when students were allowed to withhold their response, compared to when they were forced to provide an answer, demonstrating that, in a memory test, students are able to monitor which answers would be right and which would be wrong. The authors draw the same conclusion from the results of the studies on monkeys and the dolphin, results that show a pattern similar to that observed in humans.

Can we conclude that metacognitive processes are present in higher-order mammals besides humans? I believe that the answer is yes. In fact, it is common observation that in difficult tasks animals may show hesitations, signs of restlessness, and the like. As the authors mention, "these hesitations and waverings . . . made behaviorists uncomfortable because they suggested that animals might be in mental turmoil over difficult trials" (target article, sect. 4, para. 4). The cleverness of the studies reported in the Smith et al. article is to show that some of these behaviors, when related to cognitively difficult tasks, can indicate the presence of metacognitive processes. The method devised by the authors to test monitoring is also a clever modification of psychophysical tasks that have been used to test the sensory/perceptual systems in animals (e.g., Laursen & Rasmussen 1975). As such, it allows for a comparison across various species of animals, humans included.

Perhaps new studies can expand on this method. One possibility is to vary the cost/benefit ratio of the response (e.g., increasing/decreasing the cost of the error and increasing/decreasing the benefit of correct responses). This manipulation might shed light on two issues. On the one hand, it could establish whether monkeys are able to improve their performance as humans do in similar situations, by becoming increasingly selective and choosing the tasks that they want to avoid when the cost of the error is increased. On the other hand, providing a stronger drive or a

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

stronger punishment might increase the likelihood of observing the presence of metacognitive skills in less cognitively sophisticated species (if they possess any). Another possible extension of the monitoring studies might be to explore the "feeling of knowing" in animals, by measuring the their "willingness" to "search in memory" for difficult- versus easy-to-access items that cannot be immediately retrieved, and then comparing animal and human performance.

Although Smith et al.'s interpretation of the data as indicating metacognition in some animals is convincing, their assumption that metacognitive processes imply awareness is unwarranted. Indeed, some authors have claimed that many metacognitive tasks, including monitoring, are driven by implicit processes (see Reder & Shunn 1996 and some chapters in Reder 1996). Even those who believe that human metacognitive judgments are inferential note that only some of these judgments are "information-based (i.e., mediated by deliberate, analytic inferences . . .)" (Koriat & Levy-Sadot 2001, p. 34); the others are "experience-based, entailing the implicit application of global heuristics . . . These heuristics may operate below full consciousness to influence and shape subjective experience" (Koriat & Levy-Sadot 2001, pp. 34–35). The same might be true for animals as well. This does not detract from the results obtained by the present authors, nor is it in disagreement with the signal detection model that they propose and that seemingly fits the experimental data well. However, it questions some of the assumptions on which the studies are based and consequently some of the interpretations of the data. The problem is particularly evident in the claim by Smith et al. (summarized in the Abstract, but present throughout the target article) that "this exploration could extend the study of animal self-awareness and establish the relationship of self-awareness to other-awareness." This claim is not warranted, either by the results of any of the studies reported in the article, nor by the state of knowledge available at the moment about metacognitive processes in humans.

Drawing the line on metacognition

Janet Metcalfe

Department of Psychology, Columbia University, New York, NY 10027.
Metcalfe@Columbia.edu
http://www.columbia.edu/cu/psychology/metcalfe/jm.html

Abstract: Only two of the many experiments described by Smith et al., as indicating metacognitive ability in nonhuman animals, involved metacognition as understood in the human literature. Of these, one gave negative results. In the other, one of two rhesus monkeys provided data suggesting that he might have metacognitive ability. The conjecture that any nonhuman animals have metacognitive ability is, therefore, tenuous.

If metacognition means "knowing about knowing," or "reflecting on cognition," as has traditionally been assumed (see Terrace & Metcalfe, in press), then almost none of the research described by Smith et al.'s target article qualifies. The paradigm that they explore most intensively is a perceptual discrimination task. Their primary finding is that some animals can use an "uncertainty" response as well as "same" and "different" responses. This speaks to the complexity of the perceptual judgments, but it is not a finding about metacognition. Similarly, Smith et al.'s memory study (sect. 9) is a simple recognition experiment including no meta level. Human recognition is routinely conducted by asking people to use a graded scale rather than making old/new judgments. But it is not, therefore, considered to be metarecognition. Smith et al. have proposed that signal detection theory could be used to explain their tasks (consistent with 40 years of human cognitive research). However, neither a three-part decision scale nor signal detection theory can change perceptual discrimination or basic-level recognition into metacognition. To verify that I do not hold an idiosyncratic view, I looked up signal detection in 15 standard texts on human cognition, of which seven had lengthy descriptions, usually of tasks similar to those of Smith et al. None made any mention of metacognition in this context.

The fact that human, monkey, and dolphin patterns of data were similar, and that these patterns were different from those of rats or pigeons, does not mean that the behavior under consideration was metacognition. People do things other than self-reflect! The responses of primates and dolphins were more complex than those of rats or pigeons – who failed to use the "uncertainty" option – and may provide insights into the evolution of judgment processes. But these patterns do not indicate that *any* of the animals were engaged in metacognition.

Indeed, the only true metacognition in these tasks came from Smith et al.'s post-experimental interviews, in which the human participants said that they thought that "their sparse and dense responses were cued by the objective stimulus conditions"; "their Uncertain responses were prompted by personal feelings of uncertainty, doubt, and of not knowing the correct answer" or that using the uncertainty option was "a cop-out" (sect. 6, para. 3). People were clearly able to ruminate about their cognition. Their commentary applied equally to the sparse, dense, and uncertain responses – there was no special metacognitive priority given to response type. But although people are capable of this kind of metacognition, these human introspections do not suggest that animals can do the same.

Perhaps, however, animals could comment on their cognition, if only they were trained to provide a judgment about a just-made cognitive response. Smith et al. note that researchers as early as Jastrow (1888) and Brown (1910) had recommended such a procedure, but, "the catch is that animals have so far not proved able to report their confidence in this way" (sect. 5, para. 3).

Where does one draw the line distinguishing metacognitive from non-metacognitive? Most important is the prerequisite that the entity about which the judgment is made not be present in the perceptible stimulus environment. This is necessary, especially in animal research, to preclude the possibility that the animal could make the judgment about the external stimulus rather than the mental event. Metacognition must be about a retrieved representation, a mental event, or some cognitive entity, not about an external stimulus. Thus, not all judgments are metacognitive judgments. Perhaps the best-known judgment in the literature on humans that is truly metacognitive, is the feeling-of-knowing judgment. People are asked questions such as, "Who is the prime minister of Canada?" If they cannot overtly give the response, they are asked the metacognitive question: "How likely are you to remember the answer later when I ask you a 7-alternative forced-choice recognition test?" (which is given at some later time). People can make this prediction with considerable accuracy. Confidence judgments, when they are metacognitive, are about people's confidence in the retrieved response, that is, about the results of a cognitive process. They are not the response itself, however. In contrast, recognition judgments on a graded "confidence" scale, given in the presence of the probe item, are not considered to be metacognitive judgments, but rather are just graded recognition judgments – at the basic, not the metacognitive, level.

Were any of the experiments given in Smith et al.'s review metacognitive? Using the above criteria, two of the experiments were. First, in Hampton's (2001) task (sect. 10), monkeys were given a delayed-match-to-sample task. After the stimulus itself had been removed and before the memory test was given, the animal was trained to indicate whether he wanted to take the test or not. This seems similar to the feeling-of-knowing judgment just mentioned. The monkeys increased their probability of declining to take the test as the retention interval increased, but, as Smith et al. noted, this could have occurred because they associated long delays with more timeouts, and so it cannot be interpreted unambivalently as metacognitive. However, one of two monkeys showed better performance on the nondeclined trials than when forced to take the test on all trials. This finding suggests that the monkey might have been declining selectively, based on reflection

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

about his state of knowledge when he made the metacognitive (decline or no-decline) judgment.

In the other study by Shields (1999) (target article, sect. 11), monkeys associated two polygons, for a cued-recognition test. The metacognition came in when, prior to the test, they chose whether to take the test or not by pointing to either the cue polygon (to take it) or a star (to decline). Notably, the target about which the metacognitive judgment was being made was not present. Disappointingly, performance did not differ depending on their choice, suggesting that the judgments were made randomly. Only when the entire cue-target pair was presented (functionally converting the metacognitive task into a non-metacognitive recognition-memory task) did the pretest response have predictive value. This second study, then, provided no evidence of metacognition. In summary, then, the entire burden of the conjecture that nonhuman animals are capable of metacognition rests on the thin shoulders of a single rhesus monkey.

Relevance of unjustified strong assumptions when utilizing signal detection theory

Thomas O. Nelson

Psychology Department, University of Maryland, College Park, MD 20742.
tnelson@glue.umd.edu
http://www.bsos.umd.edu/psyc/faculty/tnelson/index.html

Abstract: Several conclusions depend on a version of signal detection theory that assumes performance is based on underlying equal-variance normal distributions of trace strength. Such conclusions are questionable without empirical justification for that assumption. A thought experiment is presented to show how the assumption is probably invalid, and empirical evidence is cited for the assumption's invalidity in research on human memory.

The target article by Smith et al. is valuable for extending to nonhuman animals some ideas from the literature about research on human metacognition. I agree with Smith et al. that many of their findings are difficult to explain without metacognitive mechanisms.

However, throughout their article (e.g., sect. 13, including Fig. 9A; sect. 13.2, including Fig. 10A; sect. 13.3, including Fig. 11; Appendix 1, including the simulations of the data from Shields et al. 1997; Smith et al. 1998; Hampton 2001), several conclusions drawn from Smith et al.'s "unifying formal perspective" require strong assumptions that are not empirically justified by the to-be-explained data. Specifically, their use of what has been called the "Strong Version" of signal detection theory (hereafter, SDT) assumes that what underlie performance are normal distributions (of "trace strength") that have the same variance. This assumption is problematic when such distributions are only hypothetical (cf. Lockhart & Murdock 1970) and without empirical justification in any of the to-be-explained experiments. (*Note:* This is in contrast to the "General Framework" of SDT wherein assumptions of equal-variance normal distributions are not made – the distinction is elaborated in Nelson 1987.)

Elsewhere, that assumption is empirically assessed, such as in the area of perception where the investigator manipulates the payoff matrix for errors/successes (when the participant makes responses to indicate a trace strength below/above a decision criterion). Then a graph can show a Receiver Operating Characteristic curve (ROC) in which the proportion of "old" responses to old items, Pr("old"|old), is plotted as a function of the proportion of "old" responses to new items, Pr("old"|new), on normal-normal coordinates. The main diagonal of such a graph represents nil detection accuracy insofar as Pr("old"|old) is no greater than Pr("old"|new).

In the figures from Smith et al., the leftmost distribution corresponds to the distribution of trace strength for nonpresented items (usually referred to as the noise distribution), and the rightmost distribution corresponds to the distribution of trace strength for presented items (usually referred to as the signal+noise distribution). *If both of those underlying distributions are normal, then the ROC will be a straight line,* and *if those underlying distributions have the same variance, then the slope (hereafter, s) of the ROC will be 1* because $s = \sigma_n/\sigma_{sn}$, where σ_n is the standard deviation of the noise distribution, and σ_{sn} is the standard deviation of the signal+noise distribution. Then the degree of detection accuracy is designated as d' and is computed as the distance between the *empirically obtained ROC* and the main diagonal. Most importantly, when $s = 1$ (i.e., when $\sigma_n = \sigma_{sn}$), this distance is constant regardless of where along the main diagonal the measurement occurs. However, when $s \neq 1$, the degree of detection accuracy will vary depending on where along the main diagonal the measurement occurs; then d' is not meaningful because the Strong Version of SDT is inappropriate ("according to detection theory, the index of sensitivity should remain invariant with changes in the decision criteria," Green & Swets 1966/1974, p. 110). This presents a problem for using the Strong Version of SDT in the experiments analyzed by Smith et al. because they don't contain empirical evidence about the hypothetical distributions assumed to underlie performance.

Empirical evidence that s < 1 and that $\sigma_n < \sigma_{sn}$. Contrary to Smith et al.'s implicit assumption that $s = 1$ and that $\sigma_n = \sigma_{sn}$, recent research (e.g., Ratcliff et al. 1992) has empirically confirmed in various situations that $s \cong .8$ and $\sigma_n < \sigma_{sn}$.

A thought experiment. The following thought experiment might allow the abovementioned findings of Ratcliff et al. (1992) to be made more intuitive, and might also allow a guess to be made as to what the outcome would have been if an empirical assessment had been made of the distributions that Smith et al. assumed to underlie performance in their experiments. Imagine a noise distribution having σ_n and then imagine that independently for each item in that distribution, an increment of trace strength is added by a study trial, so as to produce the signal+noise distribution having σ_{sn}. When will $\sigma_n = \sigma_{sn}$, so that $s = 1$? The answer is, *only when the increment of trace strength from the study trial is identical for every item.* Put differently, when the increment from acquisition and the original amount of trace strength are independent (as assumed by the Strong Version of SDT), the complete equation for the above analysis is

$$\sigma_{sn}^2 = \sigma_n^2 + \sigma_s^2 \qquad \text{[Eq. 1]}$$

where σ_s^2 is the variance in the signals (i.e., the variance in the increment of trace strength across all of the signals). Then by Eq. 1 the *only way in which* $\sigma_{sn}^2 = \sigma_n^2$ *is when* $\sigma_s^2 = 0$ (i.e., when the increment in trace strength is identical for every signal). However, σ_s^2 will never be zero when trace strength is a continuous variable, and even if trace strength were a discrete variable (with some reasonable degree of fineness), every item is unlikely to be incremented by *exactly* the same amount of trace strength by the study trial. Further, because σ_s^2 is a squared quantity, any variation in the increments of trace strength will necessarily yield $\sigma_s^2 > 0$. This in turn (by Eq. 1) will cause $\sigma_{sn}^2 > \sigma_n^2$ and will constitute a violation of the equal-variance assumption that is part of the Strong Version of SDT, which was the basis of Smith et al.'s simulations and "unifying formal perspective."

Thus, Smith et al.'s simulations and "unifying formal perspective" are questionable until empirical justification occurs for the assumption of normal distributions in which $\sigma_{sn}^2 = \sigma_n^2$. One possibility for obtaining ROCs from research on animals is to manipulate the payoff matrix (cf. Alsop 1998) so that each subject performs under various contingencies of reinforcement (e.g., under different penalties for making an Uncertain response), so as to assess the possibility of $s = 1$ (and assess rather than assume $\sigma_n = \sigma_{sn}$) before invoking the Strong Version of SDT.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Does metacognition necessarily involve metarepresentation?

Joëlle Proust

Institut Jean-Nicod CNRS (Ecole des Hautes Etudes en Sciences Sociales, Ecole Normale Supérieure), 75007 Paris, France; and Max-Planck Institut Für Psychologische Forschung, Münich, Germany. **jproust@ehess.fr**
http://joelle.proust.free.fr

Abstract: Against the view that metacognition is a capacity that *parallels* theory of mind, it is argued that metacognition need involve neither metarepresentation nor semantic forms of reflexivity, but only process-reflexivity, through which a task-specific system monitors its own internal feedback by using quantitative cues. Metacognitive activities, however, may be redescribed in metarepresentational, mentalistic terms in species endowed with a theory of mind.

An important conceptual issue raised by the target article consists in the sense of "self-knowledge" engaged in confidence judgments that monkeys and dolphins seem to be able to form, in contrast with other species such as rats and pigeons. The authors tend to consider that the same notion of "self-reflexivity" applies in the realms of higher mental-state attribution and of metacognitive monitoring of the system's epistemic states. Although they accept the view that cognitive self-awareness may be different from self-recognition, they suggest that metacognition is a capacity that *parallels* theory of mind: The latter asks "whether animals know and monitor the other's mental states and states of knowing," the former "whether animals know and monitor their own mental states and states of knowing" (sect. 3). This parallel may be misleading, however, in important ways. Given monkeys' lack of theory of mind and absence of self-recognition (Cheney & Seyfarth 1990; Anderson & Gallup 1997), it is highly implausible that they have any mentalistic understanding of their ability to evaluate their own epistemic dispositions. But maybe Smith et al. rather wish to suggest that metacognitive abilities constitute a precursor for the mentalistic abilities as found in humans. In this case, however, a clear distinction between mentalizing and metacognitive capacities is still needed.

Whereas by definition an animal endowed with a "theory of mind" capacity is able to monitor and predict the behavior of others in a mentalistic way, that is, by attributing mental states to others, rather than on the basis of behavioral cues, metacognition does not seem to require any mentalistic attribution, still less so "to oneself." Such a capacity presupposes that a control system (1) has access to information concerning its present epistemic states (information that "self-monitoring" provides) and (2) uses it to select and complete a particular course of action. As the authors convincingly show in the specific case of SDT ideal strategies, such a hierarchical organization optimizes the benefit/cost ratio by applying hard-wired heuristics to the endogeneous feedback. It clearly is a *procedural* form of metacognition, a "know-how to decide," that is not based on mental concepts and does not *need* to be made explicit ; even if one grants Shiffrin and Schneider (1977) that consciousness may be favorable to control, the proof of its being necessary is not made yet (see Bargh 1997; Reder & Schunn 1996; Spehn & Reder 2000).

On the background of the uncontroversial control/monitoring model, two claims made by Smith et al. deserve discussion: that metacognition is metarepresentational, and that it is "about the self."

(A) Any control system involves a form of reflexivity at the task level: There must be, as the authors write, a connection between the judgment of certainty and "the primary discriminatory process" in which it originates (cf. sects. 5, 6, 14.3, 14.4). But they add the following comment on the relation between the two "The uncertain response . . . *is about* the status of the primary discriminatory process and about its probable failure. It stands structurally *outside* the primary discrimination and *intrinsically meta* to it." (sect. 14.4, emphasis added). This observation, however, conflates "being about" and "being meta." "Being about" involves mental reference to an object, an event, or a property; "being meta" just involves hierarchical control between processes. The latter could qualify as metarepresentational if control processes modeled not only the current epistemic states of the system, but also the attitudinal contents of the latter. But why should the control system need do this? It is much more economical to have a mechanism that simply correlates the feasibility (probable success) of a task with preselected types of cues (like the quantity or intensity of the feedback), rather than one relying on the semantic processing of the first-order content of its epistemic states. Evidence of fractionation of the control system in task-specific frontal lobe modules (Shallice & Burgess 1991) is compatible with the view that no metarepresentation is taking place. The various primary processes present invariant properties that are reliably predicting success or failure in performing the corresponding task. This correlation becomes exploited when the corresponding mechanism is established by evolutionary selection and fine-tuned by learning.

(B) The notion of reflexivity at work is thus not necessarily intentional (i.e., representational) or referential; it may more plausibly be considered executive, architecture-bound or structural. The output of the control process depends in major part on the feedback it receives from the courses of action in their first-order "simulated" or "attempted" runs. But this dependence does not *need* to be semantic. An interesting "accessibility model" of how the search process is reflexively used in control is offered in Koriat (1993). Here, too, accessibility heuristics does not rely on content but on the properties of the content *vehicles* – for example, trace strength. Thus metacognition necessarily involves neither self-reflexivity (in the sense of using an integrated representation of one's own mental, social and physical dispositions) nor even mental-state reflexivity, but process-reflexivity.

This leaves us with the question of how such an epistemically implicit control system can be a step toward theory of mind (and to consciousness). Having a procedural form of metacognition puts an organism in a position to gain the corresponding form of declarative knowledge if the conditions for demodularization are met. Karmiloff-Smith (1992) hypothesizes a mechanism of "representational redescription" making knowledge contained *in* the mind accessible *to* the mind. Such ideas have since then been explored in the evolutionary history of theory of mind (Povinelli 2000). In this perspective, a metacognitive control system is a phylogenetic precursor for mentalizing ability, not only because it offers procedural knowledge to a potential redescription mechanism, but also because the resulting enhancement of executive capacities offers the control structure that decoupling requires. Inhibiting one's own view of how things look in appreciating another person's perspective, depends on adequate mental control (Perner 1998). The difference between an implicit, nonmentalistic form of metacognition and its "redescribed" or explicit form, is that reflexivity occurs not only at the process level, but also at the semantic-intentional level (Proust 2001 ; submitted). Metacognition now can be accessed by metarepresentations, and through language it becomes available to self- and other-report, to training, and, here we are, to theorizing.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Uncertainty monitoring may promote emergents

Duane M. Rumbaugh[a], Michael J. Beran[b], and James L. Pate[c]

[a]*Language Research Center and Department of Psychology, Georgia State University, Decatur, GA 30034;* [b]*Language Research Center, Georgia State University, Decatur, GA 30034;* [c]*Department of Psychology, Georgia State University, Atlanta, GA 30303.* **drumbaug@aol.com mjberan@yahoo.com jpate@gsu.edu**

Abstract: We suggest that the phenomenon of uncertainty monitoring in nonhuman animals contributes richly to the conception of nonhuman animals' self-monitoring. We propose that uncertainty may play a role in the emergence of new forms of behavior that are adaptive. We recommend that Smith et al. determine the extent to which the uncertain response transfers immediately to other test paradigms.

Smith et al. have offered those who do research with nonhuman animals an invaluable tool: a method for ascertaining not just what types of information subjects learn and retain, but also a measure of how certain those subjects are about their knowledge state. Despite the contribution to the understanding of uncertainty monitoring in nonhuman organisms, there are some problems with the current article, and we will discuss a few of these before noting other contributions of the article. Clearly, the proportions of trials on which the animals touched the box, the star, and the S with the cursor, as depicted in Figure 3 (target article, sect. 6) and elsewhere, are correct within errors of measurement. However, it should be noted that the labels on the graphs are "Sparse," "Dense," and "Uncertain" rather than "box," "star," and "S." Although the facts may be clear, they become "interpreted facts" (cf. sect. 14.2) with great ease and without apparent recognition.

The interpretation of the responses is a major problem throughout the manuscript in that the authors refer to the star response as an uncertain response in some places and as declining the trial in other places. This problem is associated with another problem if the "declining the trial" interpretation is given. Specifically, choosing the star, declining the trial, would seem to be a response that should be paired with choosing some other symbol to indicate that the trial is accepted. In particular, this is an implicit two-stage decision situation. In the first stage, the organism either declines the trial or accepts the trial. If the trial is accepted, then there is a second stage in which the organism indicates that the stimulus is dense or sparse. The analysis of a two-stage decision-making situation is different from an analysis of a single-stage decision-making situation with three alternatives. If the star indicates uncertainty, then the situation is a one-stage decision situation with three alternatives (uncertain, sparse, dense). Given the emphasis on uncertainty monitoring, it can be argued that the three-alternative interpretation is the appropriate one, and that the "declining the trial" interpretation is not germane to the issue of uncertainty monitoring.

We propose that the uncertainty response exhibited by nonhuman animals may be important in what one of us (Rumbaugh) has called emergents (Rumbaugh et al. 1996b; see also Rumbaugh 2002; Rumbaugh et al. 1996a). Emergent behaviors are new patterns of responding with no antecedent in previously learned behavior. Emergent behaviors are applied appropriately to novel situations. Perhaps uncertainty monitoring may lead to the production of emergents. Emergents may occur at moments of uncertainty when what has worked in the past will not work in the present. As noted by Smith et al., uncertainty often promotes hesitation, and we are struck by the notion that what previously has been called insight, which may be a subclass of emergents, often is the outcome of such behavioral hesitation. For example, Köhler (1925) described chimpanzees' attempts to attain out-of-reach foods before stopping, seemingly reassessing the situation, and then arriving at the use of objects to reach those items. One could imagine that the animals were uncertain about how to obtain the foods, but they recognized that a correct solution must have been available. Importantly, Rumbaugh et al. (1996b) noted that emergents "generalize between contexts not on the basis of the specific stimulus dimension, as in stimulus generalization, but rather on the basis of relations between stimuli and/or rules" (p. 59).

This notion of the generalized use of appropriate responses is an aspect of the Smith et al. uncertainty-monitoring paradigm not yet established. By this we mean the following: Human beings' subjective states of uncertainty are similar across situations. For example, when we say that we are uncertain about a person's name, we mean almost exactly the same thing as when we say we are uncertain of the exact time of day or the location of a given city. In each of these cases, we know there is a correct answer, but we also know that the answer that we would produce may not be the correct answer. As such, the feeling of uncertainty is consistent across situations. Although there may be differing levels of uncertainty, we do not qualitatively redefine our feelings across the above situations. Our question is whether nonhuman animals would use the uncertain response on a variety of transfer tasks to demonstrate that the response truly maps onto the same psychological state *from the outset.* Such transfer tests, in fact, would demonstrate consistency across such objectively uncertain states as could be produced by these tasks, but as yet, this is an unanswered question.

We agree with Smith et al. that there is no reason to assume that the use of an uncertain response by nonhuman animals is not consistent psychologically with the use of the same response by human participants. We also agree that when the objective state of the world and the subjective state of the organism coincide sufficiently, the organism relies on learned behavior. If, however, the correspondence of those states is low, the organism may produce novel responses to cope with the situation, and those novel responses may be emergents. Whether uncertainty monitoring provides information about consciousness, or working consciousness (to use the Smith et al. term), is itself uncertain. But, we suggest (along with Smith et al.) that the best possible description of uncertainty monitoring should be based on the high level of behavioral similarity between humans and nonhuman animals demonstrated in these exciting studies.

ACKNOWLEDGMENT

Support for the writing of this commentary was provided by National Institutes of Health Grant HD-38051.

Animal metacognition? It's all in the methods

Sara J. Shettleworth and Jennifer E. Sutton

Department of Psychology, University of Toronto, Toronto, ON, M5S 3G3, Canada. **shettle@psych.utoronto.ca sutton@psych.utoronto.ca http://psych.utoronto.ca/~shettle/ http://psych.utoronto.ca/~sutton/**

Abstract: When animals choose between completing a cognitive task and "escaping," proper interpretation of their behavior depends crucially on methodological details, including how forced and freely chosen tests are mixed and whether appropriate transfer tests are administered. But no matter how rigorous the test, it is impossible to go beyond functional similarity between human and nonhuman behaviors to certainty about human-like consciousness.

Devising nonverbal tests for processes normally accessed by verbal report of conscious awareness in humans is one of the biggest challenges in contemporary research on comparative cognition (Shettleworth 1998). It is one of the biggest sources of controversy as well. Many of the issues in the study of metacognition are also evident in research on whether nonhuman animals have episodic memory (Clayton et al. 2001), theory of mind (Heyes 1998), or are capable of intentional deception (Kummer et al. 1996). The challenge is to devise experimental procedures to elicit behavior from animals that is functionally similar to behavior accompanied by distinctive mental states in humans. The term *functional similar-*

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

ity (Hampton 2001) captures the idea that the best we can do in such investigations is to define rigorously the behavior accompanied by a given mental process and see if the animals show it. In general, functional similarity demands a constellation of behaviors that, together, are uniquely consistent with the process of interest. A single behavioral test is rarely enough.

Unlike Smith et al., we believe it impossible ever to go beyond a demonstration of functional similarity to infer human-like consciousness in animals. Moreover, too frequent use of anthropomorphic terms like "construal" and "uncertain" to interpret results may obscure methodological weaknesses. The history of research on related problems like theory of mind (Heyes 1998) shows that progress is most often made through dialogue between sharp-eyed behaviorist skeptics and defenders of animals' higher cognitive abilities. In this spirit of constructive dialogue, we point out some important methodological issues that are insufficiently emphasized in the target article.

Demonstrating that animals monitor memory strength or perceptual certainty requires more than showing that they choose an "uncertain" or "escape" option more often on difficult trials than easy ones. As Smith et al. acknowledge, such behavior could result from learning the relative reward rates for opting out versus taking the test in the presence of specific external stimuli. Consistent use of the escape option with new stimuli or under new testing conditions is crucial to help rule out the possibility that performance is based on specific contingency learning. In addition, to be sure that animals are assessing their cognitive state from trial to trial, "forced" tests must be interspersed randomly throughout the experiment. Mixing them in (Hampton 2001; Inman & Shettleworth 1999) rather than making these trials without the escape option relatively rare, as in some of the studies by Smith and colleagues, is essential. Worse performance on forced instead of freely chosen tests is predicted from metacognition, but animals often perform worse than normal under novel conditions.

The point in the trial at which the animal chooses to escape is also crucial. In some studies, the escape option has been presented *simultaneously* with the test stimuli, and in others, it has been presented *before* the test stimuli. Intuitively, displaying test items and asking, "Do you recognize one of these, or would you rather not answer?" tests memory strength in a less demanding way than asking, "Will you recognize the sample when you see it?" The latter tests have been used only in one study with monkeys (Hampton 2001) and two with pigeons (Inman & Shettleworth 1999, Experiment 2; Sutton & Shettleworth, in preparation). Inman and Shettleworth's findings suggest that some species may be able to discriminate memory trace strengths but not pass a more difficult test of memory monitoring.

To bolster their argument that the same psychological mechanism may underlie human and animal metacognitive performance, Smith et al. show how models based on signal detection theory can account for data consistent with metacognition. However, there is no need to assume that setting criteria and making decisions are conscious (McMillan & Creelman 1991, p. 52). Pigeons and other nonprimate animals have been widely tested in psychophysical procedures and the results successfully described by signal detection models without such implications being drawn (Commons et al. 1991). Consistent with this literature, Sole et al. (in press) recently tested pigeons in a perceptual classification task with displays of pixels on a touchscreen. Like the monkeys tested similarly by Smith et al. (in press), the pigeons most often chose an "uncertain response" with intermediate pixel densities. However, in a control not used for the monkeys, they performed no better on trials they chose to complete than on frequent random forced classification trials. Importantly, even though they did not pass this test of metacognition, all the pigeons' data could be fit extremely well by a signal detection model that assumed the birds were always maximizing the perceived reward in a consistent way.

Although Smith et al. provide a taxonomy of metacognition, researchers using animals have tended to lump together tests of perceptual certainty with tests of metamemory as if they assay a single cognitive process. But perceptual certainty ("Do I know what I am seeing?") and memory monitoring ("Do I remember what I saw?") are not necessarily the same, even though people and other animals might report on both by "saying" they are uncertain. Some species might show one and not the other. For example, pigeons passed one easy test of metamemory (Inman & Shettleworth 1999, Experiment 1), but failed a parallel test of perceptual certainty (Sole et al., in press). Future studies might more carefully distinguish between perceptual certainty and memory-monitoring abilities.

All studies discussed in the target article involve training animals very extensively to give an explicit report on memory strength or perceptual certainty. However, in everyday life people use metacognition spontaneously and implicitly, as in knowing whether to look in the phone book before making a call. Future research should explore tests related to ways animals might spontaneously use metacognition in biologically relevant contexts. For example, Call and Carpenter (2001) reported that when a goal was concealed in a simple foraging task, chimpanzees, orangutans, and children behaved as if they knew they did not know where it was. This is a promising approach that might be developed further and adapted to other species.

Metaknowledge may or may not facilitate knowledge and performance

Charles P. Shimp

Department of Psychology, University of Utah, Salt Lake City, UT 84112.
charlie.shimp@psych.utah.edu
http://www.psych.utah.edu/shimp/shimp.html

Abstract: Metaknowledge may not always facilitate acquisition of knowledge or performance of complex tasks. A pigeon, for example, depending on the task, can report what it is doing even if it cannot perform the task well, and it can fail to report what it is doing when it performs the task well (Shimp 1982; 1983).

This is a wonderfully provocative paper because it is hard to imagine anything more central to the development of a science of comparative cognition and allied fields, such as philosophy of mind, the evolution of cognition, and the neurobiology of cognition, than clarifying what nonhuman animals know about what they are doing, what they remember about what they have done in the past, and what they know about what they might do in the future – in short, what they know about their own behavior, past, present, and future. In the literature on human cognition, the past few decades have seen progress through clarifying differences between various memory systems, such as implicit versus explicit memory, episodic versus semantic memory, cognition versus metacognition, and so on. All these memory classification systems involve, in one way or another, either participants in some sense knowing that they do or do not know, or knowing that they do or do not remember. The target article by Smith et al. helps to clarify all these classification schemes by contributing to our understanding of the function of consciousness: They suggest that in at least their monitoring task, there is a sense in which knowing what you are doing – specifically, knowing one's momentary level of confidence – permits a participant to more adaptively choose stimuli.

We do not yet have an adequately general theory of nonhuman animal memory systems, and it therefore would be easy, in my opinion, to miss what I consider to be an important contribution of this demonstration. The face-value result, that a nonhuman animal can monitor its level of confidence, is important, but the demonstration of what it might mean that an animal does or does not know what it is doing, or can or cannot report how confident it is, might be even more important. In the absence of adequate theory, we must rely to some extent on plain English and intuition, and by my personal standards, the target paper by Smith et al. re-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

ports one of the most sophisticated procedures yet developed by which we can study what seems by most normal intuitive standards to be knowledge about an animal's own knowledge.

It is also important to note, I believe, that the opinion I have just expressed has a subjective component. I have suggested (Shimp, in press) that a scientific-peer-review analog of truth in advertising might be facilitated if reviewers acknowledged how their methodological and theoretical priorities might affect how they see a paper they evaluate. In that spirit, I should note that it is relevant that I already believe the distinction between knowledge and metaknowledge is an important one to clarify, and I am inclined to believe that new empirical methods that produce systematic, quantitative functional relations contribute greatly to the development of explanatory theory.

For these reasons, I see the target article in an extremely favorable light. I do suspect, however, that the authors put the adaptive function of consciousness in a more favorable light than it might generally deserve. Perhaps this is because the authors wish to focus more on the existence of metacognition in nonhuman animals than on its general function. For whatever reason, the authors emphasize the fact that animals can be better off knowing what they are doing. Presumably, the behavior of scientists is another such case: We are presumably better off being aware of the theoretical assumptions we make because, again presumably, we can in that case better evaluate them. However, it also seems clear that in some cases, such as with a performing musician, or a baseball player swinging a bat at a fastball, we are better off in many ways not knowing what we are doing, because the processing of metaknowledge might interfere with performance. In short, the authors address the case where something akin to awareness might be functionally adaptive, but there is also the case where metacognition might be correlated with poorer performance.

It has been suggested that such a case can even be demonstrated in the pigeon. As a researcher who believes that the cognitive capabilities of the pigeon are generally grossly underestimated, it was a bitter pill to swallow to read that the authors are skeptical of the pigeons' capability for metaknowledge. But their acknowledgment that the end of the story for birds is not yet written was at least a mild restorative. Without having the space to go into details – and while acknowledging that in all cases of metacognition in nonhuman animals there remain the most basic issues yet to be determined, including in my own experiments – it is possible that some kind of conditional discrimination might conceivably some day explain the results (but see Anderson & Bower 1973). I would like to suggest that the diversity of demonstrations already available puts avian metacognition in a more favorable light than do the authors (Clayton et al. 2000; Shimp 1982; 1983; Zentall et al. 2001).

I strongly endorse the authors' position on the role of parsimony in comparative cognition: This role is, most ironically, extremely complicated and needs to be evaluated very carefully in each case where it is proposed as an evaluative criterion. I agree with the authors that, while parsimony has a long and fairly honorable history, it can also lead to much mischief (Shimp 1999; 2001). If, for example, parsimony were to be used to reject out of hand the kind of conceptual development the authors propose, I believe it would be sharply counterproductive.

The authors have been so successful that I am encouraged to wonder about future possible experiments using a monitoring response, particularly with avians. Can a monitoring response be successfully used in research on avian visual multidimensional categorization performances (Herbranson et al. 1999; 2002), using a method adapted from human research (Ashby & Maddox 1998)? That is, would a pigeon more frequently choose not to categorize more difficult stimuli? Finally, does the monitoring response depend on temporal parameters of a task in ways that at least intuitively correspond to the functionality of consciousness?

Implicit metacognition, explicit uncertainty, and the monitoring/control distinction in animal metacognition

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

[a]*Department of Psychology, Barnard College, New York, NY 10025;* [b]*Department of Psychology, Florida International University, Miami, FL 33199;* [c]*Department of Psychology, Columbia University, New York, NY 10027.* **lson@barnard.edu Bennett.Schwartz@fiu.edu nkornell@psych.columbia.edu http://www.fiu.edu/~schwartb http://www.columbia.edu/~nk267**

Abstract: Smith et al. demonstrate the viability of animal metacognition research. We commend their effort and suggest three avenues of research. The first concerns whether animals are explicitly aware of their metacognitive processes. The second asks whether animals have metaknowledge of their own uncertain responses. The third issue concerns the monitoring/control distinction. We suggest some ways in which these issues elucidate metacognitive processes in nonhuman animals.

In the target article by Smith et al., the authors demonstrate that nonhuman, nonverbal animals are able to inform us that they feel "uncertain," one proposed measure of metacognitive ability. Prior to these data, it was not known whether an animal could report introspective assessments about retrieved memories and response choices. This new and exciting line of research has sparked debate, discussion, and further studies investigating the level of metacognitive abilities in nonhuman species, as well as shaking the definition of metacognition itself. We commend Smith and his colleagues for daring to go where no researchers had gone before. We discuss three potential avenues for research in this new field of animal metacognition: the implicit/explicit distinction, knowledge of uncertainty, and the monitoring/control distinction.

One exciting discussion is that human metacognitive processes involve both explicit and implicit influences (Reder & Schunn 1996). Traditionally, researchers have defined metacognition as the ability to explicitly assess the certainty of retrieved items from memory and consciously verbalize the judgment. However, some recent research suggests that some metacognitive processes may act without conscious awareness (Reder 1987; Reder & Schunn 1996). For example, Reder (1987) showed that people could decide that they knew something more quickly than they could actually retrieve the information – indicating a lack of explicit awareness in the decision. If uncertainty is an implicit "feeling," animals may be able to respond similarly to how they might respond to a "feeling" elicited by a fear-invoking stimulus. Thus, some now propose that the definition of metacognition must encompass both explicit and implicit abilities to be uncertain, whether it is reported verbally or behaviorally (see Son & Kornell, in press).

A second and related issue concerns the animal's knowledge of its own uncertain responses – a genuine *meta*cognitive process. Most of the research in the Smith et al. target article relies on evidence using an opt-out response, an option simultaneously presented with other response options. When uncertain, the animal can escape negative consequences of being incorrect on that trial. This escape response cannot address the issue of whether the animal has knowledge of being uncertain, because the option is presented concurrently with other options, as well as with the features of the question (i.e., Hampton 2001) – making it more similar to a mere cognitive response, rather than a *meta*cognitive response. The escape response confounds the object-level task (e.g., which patch is denser), and the meta-level task (i.e., confident or uncertain) does not distinguish between an animal motivated to avoid penalties and one actually becoming aware of its uncertainty.

Can animals show us that they know how uncertain they are? This would entail that the animal makes a metacognitive response after a cognitive response, not simultaneously. A procedure developed by Son et al. (2002) investigated the issue of whether animals can make these higher-level metajudgments (also see Hampton 2001). Rhesus macaques were asked to choose the longest of nine lines and, once the lines disappeared from the

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

screen, they made a high- or low-confidence judgment. Results showed that the monkeys reliably chose high confidence after easy trials that were more likely to be correct and low confidence after difficult trials that were more likely to be incorrect. The animals acted metacognitively and, moreover, continued to do so when the risk task was transferred to other cognitive tasks that did not involve lines. This retrospective judgment suggests that not only is the animal motivated to avoid penalized responses, but that it can report knowledge of its state of uncertainty. Analogously, a human might feel one line is the longest, but not be sure, or be absolutely certain they were guessing. Son et al.'s task is easily modifiable to distinguish between states of low confidence and guessing.

Another issue in the human metacognitive literature is that there is an important difference between metacognitive monitoring and metacognitive control. Monitoring refers to peoples' ability to become aware of how well their cognitive processes are working. To do this, the monitoring process builds a model of how the basic process is working. Monitoring processes serve as the basis for metacognitive judgments. Control refers to a decision that the metacognitive system makes based on the output of the monitoring system. If the monitor informs the person that learning is going poorly, the control processes can redirect attention to the necessary to-be-learned items. For the control to be effective, it must be able to change the object-level processes and change them in adaptive ways. The effectiveness of control processes is partially dictated by two factors: (1) the monitoring effectiveness, and (2) the choice of good strategies.

The differences between metacognitive monitoring and control have not been explicitly addressed with respect to animal cognition. As suggested by other domains, animals may also show poor control in situations in which they have accurate knowledge (Boysen et al. 1996). We suspect that animals will not only vary in the degree to which they behave metacognitively, but in the kinds of metacognition they demonstrate as well. We also suspect that some species will excel at monitoring tasks (perhaps those frugivores who have to make fine distinctions between fruit colors), whereas others might control behavior better (e.g., those that need to outwit larger conspecifics). Like Smith et al., we endorse a broad comparative approach using a large range of species. Using foraging ecology to generate hypotheses has greatly benefited the study of spatial memory (e.g., Burke et al. 2002; Platt et al. 1996), and we believe that such investigations will also help us elucidate other issues of metacognition as well as the differences between monitoring and control. We contend that being able to differentiate between monitoring and control processes – in addition to explicit/implicit monitoring and knowledge-of-uncertainty/uncertainty responses – may help researchers define the boundaries of metacognitive abilities, and thus is consistent with Smith et al.'s call for a comparative study of metacognition. Animal metacognition is still an open book with untold interesting questions and methodologies to be discovered.

ACKNOWLEDGMENTS

Part of the research described in this commentary was funded by NIH grant MH40462 to Herbert S. Terrace. We would like to thank the members of the Terrace Lab: Jessica Cantlon, Dustin Merritt, Tammy Moscrip, Francys Subiaul, and Herb Terrace; the Metcalfe Lab: Brady Butterfield, Bridgid Finn, and Janet Metcalfe; and the Schwartz Lab: Megan Hoffman and Isabel Sanchez.

Uncertain what uncertainty monitoring monitors

Victoria M. Wilkins, LeeAnn Cardaciotto, and Steven M. Platek

Department of Psychology, Drexel University, Philadelphia, PA 19102.
{Victoria.M.Wilkins;LeeAnn.Cardaciotto;Steven.M.Platek}@drexel.edu

Abstract: Smith et al. present a model that they suggest may clarify aspects of the phylogenetic distribution of metacognition, based on observation of what they call uncertainty monitoring. Although they suggest that their model is supported by data collected using monkeys and dolphins, their interpretation that nonhuman animal behaviors parallel thought processes in humans may be unwarranted. The model presented by Smith et al. is inconsistent with current theories and empirical findings on the comparative aspects of metacognition. We present three oversights of the model and extend our critique to include a brief discussion of animal self-awareness, as well as current neuropsychological perspectives on metacognitive processing in humans.

The model suggested by Smith et al. is interesting, but only with respect to human participants. The experimental tasks, although interpreted as paralleling human thought processes, are wrought with methodological issues when applied to nonhuman animals. The first problem (briefly acknowledged by the authors) is the fact that the comparative uncertainty responses may simply be trained; that is, the animals simply could be learning a third behavioral response, which are associated with rewards (e.g., shorter timeouts) in the data. The authors, however, interpret this behavioral response as an uncertain response. The authors appear to attribute verbally communicated human thought processes to nonverbal, nonhuman animals! Even in verbal humans this response may not be what it appears: Humans, for social desirability and lack of a complete understanding of their mental processes, may attribute undeserved meaning to what are, essentially, trained responses.

Our second contention with the paradigm is the lack of ecological validity. Although the authors state two requirements for correctly executing this type of paradigm, "create perceptual or cognitive difficulty for the animal in order to stir up something like an uncertainty state" (sect. 4) and "provide a behavioral (i.e., nonverbal) response that lets the animal comment on or cope adaptively with that state" (sect. 4), they fail to test the model under conditions that are likely to be encountered by their subjects in the wild (e.g., uncertainty with respect to monitoring dominance position, alliances, face perception). The authors assume that what creates uncertainty in humans also creates uncertainty in monkeys, but differences in environment and environment-specific functioning suggest that the stimuli of uncertainty would also be different.

Our third issue is the authors' definition and use of the term "metacognition." The authors define the term as "thinking about thinking," and although we agree with this definition, the authors fail to adequately address the vast literature investigating metacognitive capacities in nonhuman primates (e.g., Hare et al. 2000; 2001; Povinelli et al. 1993; 1998; Tomasello & Call 1998), which has consistently provided support for a model developed over a decade ago by Gallup (1982) that theory of mind (i.e., metacognition) can only emerge in organisms that can first conceive of themselves (e.g., self-awareness; Gallup 1988; Gallup et al., in press). This model has gained increasing support in the last decade (e.g., Frederick & Platek, under review; Shillito 2002; Shillito et al. 1999; see also Amsterdam 1972; Gallup et al. 2000; Perner & Wimmer 1985). The authors also fail to acknowledge that the mark test (Gallup 1970) has been well supported by empirical data (see Shillito et al. 2002) from across the world and from almost every living species of nonhuman primate. The data strongly favor the notion that only great apes can pass the mark test. By using a limited phylogenetic distribution of animals and a very low sample size, the authors fail to adequately accomplish what they set out to do – examine the comparative psychology (i.e., phylogenetic distribution) of metacognition. Furthermore, the vast liter-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

ature showing that monkeys consistently fail tests of self-recognition, visual perspective taking, deception, and metacognition (Anderson et al. 1994; Gallup 1970; Suarez & Gallup 1981; see also Gallup et al., in press; Shillito 2002 for reviews of this literature), questions the appropriateness of the authors' use of monkey data for the centerpiece of their proposal. Likewise, with respect to dolphin metacognition the data available are both limited and questionable at best (see Reiss & Marino 2001a).

We would like to highlight a study conducted by Menzel et al. (1985) that might further support the notion that *monkeys* are incapable of monitoring uncertainty states that they experience. Menzel designed a task in which rhesus monkeys and chimpanzees reached for food; however, the subjects were only given visual access to their limb through a mirror reflection. Unlike chimpanzees that immediately were able to guide their hand to the food reward, rhesus monkeys could neither obtain the food reward nor conceive of the fact that the hand in the mirror was their own hand. Indeed, Menzel reported that several of the rhesus monkeys vocalized at the mirror reflection of their hand as if it were another monkey's hand going for the reward! Thus, in a clear situation in which uncertainty monitoring must be reconciled with a self-conception and a metacognitive self-representation, monkeys, but not great apes, failed! Monkeys may be quick learners, but the spontaneity of chimpanzees in the negotiation tasks, in the absence of demonstration or modeling, suggests that chimpanzees possess something more akin to "metacognition."

Furthermore, Smith et al. lend no discussion to current neuropsychological literature describing that many metacognitive activities occur in the prefrontal cortex. The evolutionarily recent and highly intricate prefrontal cortex has been implicated in many of the processes related to metacognition, such as theory of mind (Baron-Cohen et al. 1994), self-recognition (Keenan et al. 2001; Platek & Gallup 2002), empathy (Gallup & Platek 2002; Platek et al., in press), visual perspective taking (Stuss et al. 2001; Vogeley et al. 2001), and deception monitoring (Stuss et al. 2001). When one compares prefrontal cortices of monkeys and great apes, those of the former clearly are not as sophisticated as those of the latter (Sherry 2000).

While the model presented by Smith et al. is arguably appropriate for human uncertainty monitoring, the application of this model/paradigm to nonhuman, nonverbal animals as a means with which to study the comparative psychology of metacognition appears to be less warranted.

ACKNOWLEDGMENT

The authors thank Gordon G. Gallup, Jr. for comments on an early draft of this commentary.

Evidence both for and against metacognition is insufficient

Thomas R. Zentall

Department of Psychology, University of Kentucky, Lexington, KY 40506
zentall@uky.edu
http://www.uky.edu/AS/Psychology/faculty/tzentall.html

Abstract: The authors' attempt to explore the ability of animals to monitor how certain they are of their choice behavior, necessarily fails both in their effort to include "higher" mammals (such as monkeys and dolphins) in the class of metacognitive organisms (humans) and in their conclusion that "lower" organisms are not capable of similar behavior.

In arguing for metacognition in animals, the authors assume that if the use of the Uncertain response shown by humans and monkeys is similar, and if humans describe their decision process in metacognitive terms (reporting an awareness of their decision process), then it is likely that the monkeys too, are aware of their decision processes. A more parsimonious view is that the animals have learned the various contingencies of reinforcement: If the discrimination is particularly difficult, they choose the Uncertain response when the outcome following that response (e.g., immediate replacement with an easier trial) is more favorable than the alternative choice (i.e., the weighted average of the rewarded correct choice and the punished incorrect choice that is followed by a timeout).

The authors acknowledge that "simpler" associative learning processes might be involved but they reject this possibility for four reasons (sect. 14.2). First, they claim that animals would have to learn a large number of contingencies. But animals could instead learn a simple associative rule, "choose the High or Low value when discriminability is high and the Uncertain alternative when discriminability is low."

Second, they argue that parsimony may be wrong. Sometimes the more complex theory is correct. Although they may be right, it is not a sound argument for accepting the more complex theory.

Third, they propose that it is less parsimonious to suggest that humans use metacognitive processes but animals use simpler, associative processes to produce quite similar behavior. But it certainly could be true. (Petroleum and electric powered automobiles may behave in similar ways, yet the underlying mechanisms producing those behaviors are quite different.)

Finally, the authors consider the parsimonious alternative, that the choice behavior of both humans and monkeys is governed by simpler associative processes and that "humans' introspections and reports of reflexive consciousness and metacognition are a nonfunctional epiphenomenon" (sect. 14.2). But they counter by arguing that giving up such higher-level processes in humans "is a lot to pay for reserving the right to dismiss animal minds." An unexplored possibility is that metacognition is an epiphenomenon, but a functional one. It is possible that metacognition does not precede a difficult choice but rather, follows or justifies it. The function of metacognition for humans may not be to help in making decisions but to reduce the aversiveness associated with a less than ideal outcome – punishment (for an incorrect response) or a less desirable outcome (for having chosen the Uncertain over a possible correct alternative).

The authors' second conclusion is that metacognition may be a process that is limited to primates (and dolphins). If one accepts the fact that evidence for such behavior must be nonverbal, then the distinction between nonverbal primates and other animals will be difficult to make. First, evidence that other animals do not show behavior that the authors would judge as indicative of having metacognition, depends on acceptance of the null hypothesis – surely, a precarious stance. I am reminded of the claim by Bitterman (1960) that qualitatively different performance on a task such as serial reversal learning could be used as a measure of relative intelligence among species. Bitterman reported that with repeated reversals of a simultaneous discrimination, rats and pigeons show progressive improvement in rate of acquisition, but goldfish do not. However, Mackintosh (1971) found that a minor change in the procedure can lead to similar improvement in the rate of acquisition of serial reversals for both goldfish and pigeons.

The authors note that when humans are confronted by a difficult choice, they report feelings of uncertainty. But similarly interpreted behavior has been reported in rats' "vicarious trial and error" behavior (Tolman 1938) as well as the "neurotic" behavior shown by dogs when discriminations are made difficult (Pavlov 1932).

According to the authors, the evidence against metacognition in nonprimates comes from two studies with pigeons, only one of which was published (Inman & Shettleworth 1999). But these data do not provide convincing evidence against metacognition in pigeons. The authors suggest that to qualify as metacognition, the use of an Uncertain response must show two characteristics: First, the frequency of the Uncertain response should increase as accuracy on the task decreases (one monkey tested by Smith et al. 1997 showed such an increase, the other did not; compare Figs. 4C and 4D). For pigeons, Inman and Shettleworth found a small but re-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

liable increase in use of the Uncertain response as the retention interval increased.

Second, Smith et al. suggest that one would expect performance on the task to be higher when the Uncertain response is permitted but not chosen, than when the Uncertain response is not an option. Presumably, the Uncertain response would be made on trials involving the greatest uncertainty and thus the greatest likelihood of making an error. Once again, although the effect was relatively small, the three birds that used the Uncertain response showed significantly higher matching accuracy on the task when the Uncertain response was permitted than when it was not. (Performance by the fourth bird was sufficiently high that it rarely used the Uncertain response.) Although the magnitude of the effects was not as great as that for monkeys, the results were not qualitatively different.

Thus, because, on the one hand, it is difficult to determine conditions under which even verbal humans unambiguously monitor what they know for purposes of decision making, it may be impossible to study such monitoring in nonverbal animals. On the other hand, although there may be quantitative difference in the general cognitive abilities of monkeys and other animals, there is growing evidence that many of these differences are not qualitative in nature. For example, pigeons appear quite capable of behavior that, if reported in humans, would be considered imitative learning (Zentall et al. 1996), episodic memory (Zentall et al. 2001), symbolic representation (Zentall 1998), and transitive inference (Weaver et al. 1997).

ACKNOWLEDGMENT

Preparation of this commentary was supported by Grant MH-59194 from the National Institute of Mental Health.

Authors' Response

Inaugurating a new area of comparative cognition research

J. David Smith[a], Wendy E. Shields[b], and David A. Washburn[c]

[a]*Department of Psychology and Center for Cognitive Science, State University of New York at Buffalo, Buffalo, NY 14260;* [b]*Department of Psychology, University of Montana, Missoula, MT 59812;* [c]*Department of Psychology and Language Research Center, Georgia State University, Atlanta, GA 30303.* **psysmith@acsu.buffalo.edu**
wshields@selway.umt.edu dwashburn@gsu.edu
http://wings.buffalo.edu/psychology/labs/smithlab/
http://psychweb.psy.umt.edu/faculty/shields/shields.html
http://www.gsu.edu/~wwwpsy/faculty/washburn.htm

Abstract: There was a strong consensus in the commentaries that animals' performances in metacognition paradigms indicate high-level decisional processes that cannot be explained associatively. Our response summarizes this consensus and the support for the idea that these performances demonstrate animal metacognition. We amplify the idea that there is an adaptive advantage favoring animals who can – in an immediate moment of difficulty or uncertainty – construct a decisional assemblage that lets them find an appropriate behavioral solution. A working consciousness would serve this function well. This explains why animals may have the functional equivalent of human declarative consciousness. However, like other commentators who were friendly to this equivalence, we approach carefully the stronger claims that animals' metacognitive performances imply full-blown self-awareness or phenomenal consciousness.

We discuss the commentators' interesting ideas for future research, as well as their intriguing ideas about the evolution and development of metacognition and its relation to theory of mind. We also discuss residual confusions about existing research and remaining methodological issues.

R1. Introduction

We thank you all for offering your perspective from many areas of cognitive science. We believe the give-and-take allowed by *Behavioral and Brain Sciences* serves this new area of inquiry well. Before beginning the hard work of replying in this Response, we pause to reflect on the positive. This does not always happen in a field that values critical sharpness above everything. But such reflection is helpful in seeing where the area stands.

Nelson agreed that many of the target article's findings are difficult to explain without metacognitive mechanisms. **Goldsmith & Koriat** granted the possibility that monkeys and dolphins may monitor their own state of knowledge and control their behavior accordingly. **Mazzoni** thought the data convincingly demonstrate metacognitive skills in animals. **Flavell** hypothesized, based on the target article, that animals might show conscious uncertainty responding. **Marino** thought we provided convincing evidence and a reasonable argument for declarative consciousness as a shared psychological property in humans, monkeys, and dolphins. **Son**, **Schwartz & Kornell** (henceforth **Son et al.**) acknowledged these first demonstrations that animals can inform us that they feel "uncertain," one proposed measure of metacognitive ability. **King** concluded that rhesus monkeys can apparently make accurate assessments of their own uncertainty and express that information. **Fantino** thought we made a strong case for parallels between human and nonhuman metacognition. **Griffin** placed uncertainty, survival, and consciousness in an ecological framework. He said, "If animals are aware of anything, the many uncertainties that are critical for survival must often require conscious attention."

Rumbaugh, **Beran & Pate** (henceforth **Rumbaugh et al.**) linked uncertainty monitoring to animals' highest-level (emergent) modes of cognition. They called the uncertainty-monitoring paradigm an invaluable tool. **King** noted that the uncertainty response adds an important new dimension to conventional animal learning and memory studies. **Shimp** summarized the broad potential interest in this area:

> It is hard to imagine anything more central to the development of a science of comparative cognition and allied fields, such as philosophy of mind, the evolution of cognition, and the neurobiology of cognition, than clarifying what nonhuman animals know about what they are doing, what they remember about what they have done in the past, and what they know about what they might do in the future.

He granted us one of the most sophisticated procedures yet developed for studying animals' self-knowledge.

In short, there was more positive support than we could have hoped for. Indeed, we take the overall tone of positivity and interest to mean that the existing research, the target article, the commentaries, and this response, together

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

inaugurate a new area of comparative inquiry that has substantial theoretical interest, rich connections to other areas, and endless possibilities for further research. In turn, though, many of the above-mentioned and other commentators criticized aspects of our research and our exposition. Unfortunately, we found no "uncertain or decline" option available on the *BBS* website. Therefore, we will now summarize our reflections on the commentaries.

R2. What do pigeons know and when do they know it?

There was residual confusion surrounding the pigeon results (Inman & Shettleworth 1999; Teller 1989). **Hampton**, **Shettleworth & Sutton**, and **Zentall** re-asserted that pigeons fail a strict test of metacognition but pass an easy one. We need to be sure the empirical/inferential picture is clear.

In Teller's (1989) experiment, pigeons showed a 20% increase in safe-key use at longer forgetting intervals. However, they showed almost no advantage (only 2%) for chosen over forced trials. This advantage would be larger if birds chose to complete trials based on a positive metacognitive signal. All agree that pigeons failed this test of metacognition.

In Inman and Shettleworth (Experiment 2), the birds showed almost no chosen-forced advantage (3%) and no reliable increase in safe-key use with longer forgetting intervals (5%). All agree that pigeons failed this test, too.

Inman and Shettleworth (Experiment 1) gave pigeons the choice objects simultaneously with the safe key. The results were almost identical to those in Experiment 2. The 5% (compare 3%) chosen-forced advantage was not significant. The 7% (compare 5%) increase in safe-key use with longer forgetting intervals was also not significant by a parametric test. The birds failed this test, too. One can hardly imagine two more similar results (5%–3%; 7%–5%).

The confusion arises because Inman and Shettleworth substituted a nonparametric test for a parametric test in one case in their article, and then the 7% became significant. When they restricted their focus to three birds, the chosen-forced advantage increased. Readers must decide upon the weight these secondary analyses bear.

However, these results absolutely do not support the notion that pigeons fail a strict metamemory test but pass an easy one. The strict-easy interpretation requires a significant difference *between* data patterns. This wasn't tested for, and it isn't there (5%–3%; 7%–5%). Moreover, we discuss next why the two studies probably involve the same kind of metamemory assessment. The strict-easy interpretation is thus incorrect empirically, and probably conceptually too. In the end, everyone agrees that pigeons show (so far) almost no capacity for metacognition.

R3. Simultaneous versus prospective monitoring tasks

Several commentators (**Hampton**, **Metcalfe**, **Rumbaugh et al.**, **Shettleworth & Sutton**, **Son et al.**) focused on the distinction between simultaneous and prospective judgments of knowing. The former occurs when the discrimination responses appear with the uncertainty response so that participants consider their whole situation in deciding whether to complete the trial. The latter occurs when participants make the decline-accept decision before receiving the discrimination responses.

Shettleworth & Sutton and **Hampton** thought that the simultaneous approach tests metamemory less demandingly. This idea grounded their belief that pigeons fail a difficult, prospective test of metacognition but pass an easy, simultaneous test in which they make the metacognitive judgment in the presence of the test stimuli. Section R2 above explained why this belief is poorly founded. To **Metcalfe**, a prerequisite for a metacognitive performance is that the entity being judged should not be perceptibly present. **Son et al.** suggested (without explaining why) that the simultaneous procedure might confound the object-level task (e.g., dense-sparse) and the meta-level task (confident, uncertain). **Rumbaugh et al.** seemed to favor the simultaneous task for the purpose of research on uncertainty monitoring.

On reflection, it seems there are many reasons to conclude that the simultaneous-prospective distinction is not interpretatively important and that the two tasks are empirically and psychologically the same. First, the literature uses the stimulus-available approach without devaluing it, as when participants study pairs of items in a paired-associate task and judge, while each pair is present, how well they know it or will know it later. Second, the literature even suggests that the stimulus-present metacognitive judgment is especially difficult (not easier) because the perceptual units active in working memory may dominate the monitoring or blind the metacognitive capacity (e.g., Dunlosky & Nelson 1992).

Third, it is clear that humans are metacognitive about present stimuli. When a gas gauge reads near-empty, it promotes metacognition about whether we still have enough fuel to get home before we're stranded. When an SUV approaches a command center in Iraq, it promotes metacognition about whether the vehicle holds belligerents or frightened civilians. Present stimuli do not dim or obviate metacognition; they often occasion it. Fourth, as a related example, the simultaneous-prospective distinction would imply that metacognition would occur for fill-in-the-blank questions (because, for these, uncertainty must be judged prospectively in the absence of the response options) but not for multiple-choice questions (because for these the uncertainty is judged with the response options simultaneously available). No one would endorse such a tightly restricted sphere for metacognition. Indeed, the whole idea of the show, "Who Wants to Be a Millionaire?", is to have contestants be publicly, painfully metacognitive but with the choice information simultaneously available with the decline response (take the $64,000 and run). Multiple-choice questions may even offer richer and more diverse metacognitive trains of thought than do fill-in-the-blank questions, because each of the multiple response options can promote memory search and metacognition independently.

Fifth, Smith et al. (1998) and Hampton (2001) showed highly similar decisional strategies by animals given simultaneous and prospective monitoring tasks (Figs. 10A and 11 in the target article). Sixth, Inman and Shettleworth (1999) found almost identical performance by pigeons given simultaneous and prospective procedures. Seventh, the same

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

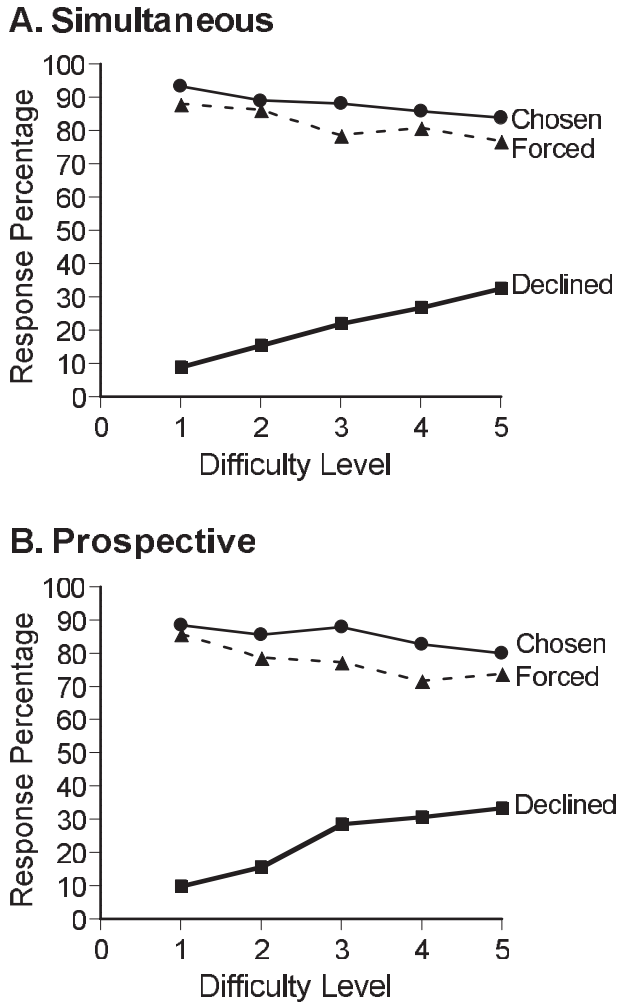


Figure R1. Humans' performance in a match-to-sample task. The stimuli were 9-point random-dot polygons presented for varying brief/tachistoscopic intervals. Stimuli were masked after offset on 50% of the trials. Simultaneous participants then saw two shapes (a foil and the just-seen sample) and a star with which they could express uncertainty. Prospective participants made an initial decline-accept decision (the star vs. a yellow dot), and then completed the trial if they elected to. Humans were given the choice to respond uncertain/decline on 80% of trials; on 20% of trials they were forced to answer. The percentages correct on chosen and forced trials are shown, as well as the percentage of trials declined at each difficulty level (i.e., at each level of sample duration).

appears to be true of humans. We gave humans a match-to-sample task in which the samples were available for differing (brief) amounts of time and sometimes masked after presentation. These manipulations varied difficulty and created uncertainty. The simultaneous participants then saw a response screen with the uncertainty response and two shapes (one just presented; one not). The prospective participants made a decline-accept decision and then (perhaps) completed the trial. Both groups also received occasional forced tests in which they had to complete the trial. Humans showed the same, ideal metacognitive data pattern in both cases. See Figure R1: The chosen-forced advantages were 6.2% and 7.6%, respectively; the mean and the increase in the percentage of declined trials were essentially identical. It would be difficult to imagine two more similar results, though Inman and Shettleworth's pigeons performed almost this similarly.

At present, there seems to be no reason to make a distinction between the simultaneous and prospective procedure.

R4. Other methodological issues

The commentators considered other strengths and weaknesses of current methods. We join this discussion.

R4.1. Forced trials

Shettleworth & Sutton recommended intermixing forced trials that the animal must complete. Performance on forced trials should blend performance on trials the participant would have accepted and declined. Thus, performance on forced trials should be lower than performance on accepted trials that presumably are accepted based on a favorable metacognitive signal.

The forced-trial approach can be useful but presents problems. First, forced trials may jar animals by changing their response repertory from trial to trial. This could reduce performance on forced trials for non-metacognitive reasons. Second, forced trials do not let one measure performance directly following weak metacognitive judgments. Third, this approach is inefficient – researchers have sometimes devoted 33% of trials to uninteresting forced-decline trials that balance the forced-accept trials in the design (Inman & Shettleworth 1999). Fourth, forced trials can yield insensitive measures. For easy trial types which the animal would mostly accept anyway, there may be too little difference between performance on forced and chosen trials to discern statistically.

As a complementary approach, Shields et al. (2003) and Son and Kornell (in press) proposed that animals be given the analog of a confidence-rating scale so that following each primary discrimination response they rate their confidence in the response. This approach lets one integrate strong and weak metacognitive judgments nonjarringly into ongoing performance and measure performance for both. It is efficient (no trials are wasted) and the response repertory never changes. In section R5 we illustrate humans' and a monkey's use of a confidence-rating scale.

R4.2. Reinforcing uncertainty responses

Some procedures reward the animal for uncertainty responses (Hampton 2001; Inman & Shettleworth 1999). Unfortunately, this approach encourages the idea that the uncertainty response is just a small-reward response that is conditioned in some stimulus contexts or that carries some general appetitive strength. This approach offers comfort to associative accounts.

There are alternative approaches. Teller (1989) let the safe key highlight the correct response. This hint made the trial easy, but the animal still completed the trial, and this was the reinforcement context. We have used the same hint/trial-completion method. A constructive methodological goal is to distance the uncertainty response from the reward grammar of the task, insofar as this is possible.

Interestingly, no one has used an uncertainty response that only ends one trial and produces the next. This shows how committed we all are to the need for reinforcement to

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

establish and control behavior. This uncertainty response would be interesting to evaluate because it would be the most distant from reinforcement.

R4.3. Small stimulus sets

Some procedures recycle the same few samples through whole sessions or experiments (three samples in Teller 1989; Inman & Shettleworth 1999; four in Hampton 2001).

This is a sturdy approach to the delayed matching-to-sample (DMTS) paradigm, but it might change the kind of memory search and metamemory process the task shows. If, in a sense, every sample is either Peter, Paul, or Mary (or, in **Hampton**'s case, John, Paul, George, or Ringo), the memory search is highly constrained. It could become automated and attentionless because the to-be-searched memory locations are so constant and so strongly primed by the task. A goal, therefore, is to expand the target sets. At the limit, the set could be infinite and nonrepeating, and this would be a valuable complementary assessment of metamemory. Then the animal would not be able to automatically search preprogrammed memory locations but would have to consider whether it remembers the novel item just shown.

R4.4. "False" psychophysical tasks

Shettleworth & Sutton, in focusing on the critical role that method plays in this area, discussed a study by Sole et al. (in press). The latter tested pigeons in a perceptual task that was meant to replicate the dense-sparse uncertainty-monitoring tasks used with humans and monkeys. The idea to generalize that task to a third species is sound. Yet this experiment shows how critical it is that one should use established techniques and replicate existing paradigms closely.

The problem is that Sole et al. did not adjust animals' thresholds dynamically, as is done in almost all psychophysical procedures. (This point also escaped **Wilkins et al.** These commentators said we assumed that what creates uncertainty in humans also creates uncertainty in monkeys. Psychophysical tasks do not make this assumption – they adjust trial difficulty relentlessly to find each creature's point of uncertainty.) Instead, Sole et al. used fixed density levels throughout testing for the difficult sparse and true dense trials. These stimulus levels differed by 14%, possibly beyond a JND. Indeed, these stimulus levels seem discriminable with no training (Fig. R2a). It is unlikely this fixed procedure found birds' thresholds or tested their uncertainty-monitoring capacities.

Illustrating this point, we ran humans with the fixed stimulus levels seen by pigeons or with threshold adjusted as in true psychophysical procedures. The fixed group (Fig. R2b) essentially never responded uncertain because they never were. The threshold group (Fig. R2c) replicated Smith et al.'s dense-sparse results. Sole et al.'s procedure is impossible to interpret because it does not reproduce the essential threshold phenomenon. It didn't do so with pigeons, either. Pigeons were either above or below chance performance (i.e., they weren't at threshold), and their use of the safe key was only about 5% more for difficult-sparse compared to true dense stimuli (for monkeys, this difference was 50%).

We believe it is crucial that re-creations of phenomena be that, and not the shadow of phenomena that cannot advance or clarify the empirical situation.

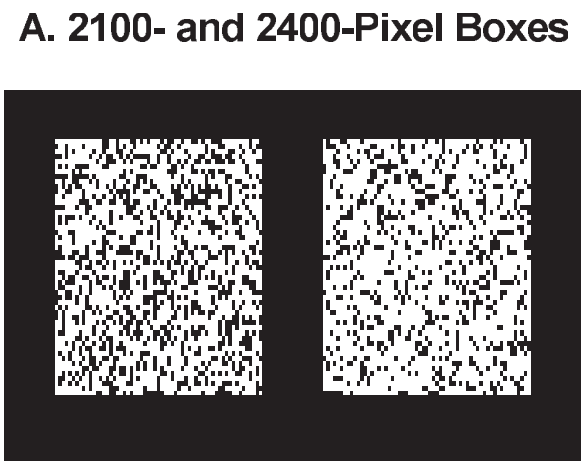


Figure R2A. Difficult sparse and true dense stimuli created following the stimulus description in Sole et al. (in press).

R5. Research directions

Colleagues envisioned many productive lines of research that we consider here.

R5.1. Generalized uncertainty

Rumbaugh et al. noted that humans make similar responses to similar feelings of uncertainty in many contexts, but that animals have been given the uncertainty response in specific tasks. Thus, research should still explore animals' flexible, general use of the uncertainty response to strengthen the inference that a general cognitive state is being monitored. For example, consider a learning-set para-

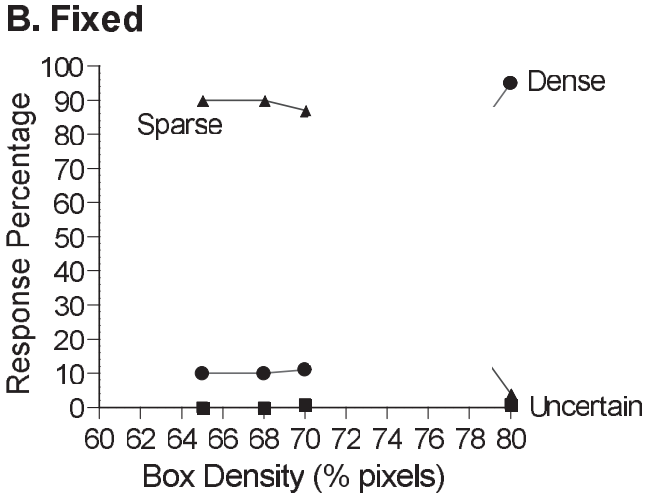


Figure R2B. Results when eight humans were run for about 900 trials each in a difficult-sparse versus true-dense discrimination with the option to respond uncertain. The difficult-sparse stimuli grew gradually denser in training then, on reaching 2100 pixels, were fixed at that level of about 70% full coverage of the box. True dense stimuli contained 2400 pixels, or about 80% full coverage of the box. The percentages of sparse, dense, and uncertain responses at each trial level are shown.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

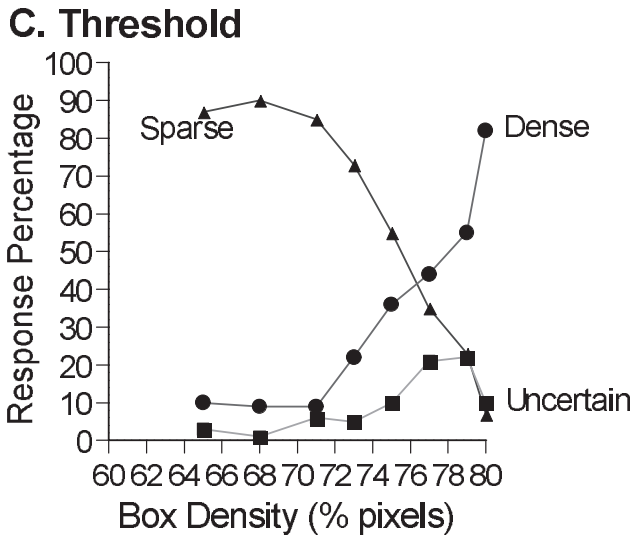


Figure R2C. Results when seven humans were run for about 700 trials each in a true psychophysical procedure in which the difficult-sparse stimuli were adjusted in difficulty dynamically based on performance to find participants' real thresholds. The percentages of sparse, dense, and uncertain responses at each trial level are shown.

digm in which animals receive a long series of two-object choice problems (with one object the S+). Normally, on the first trial of a new problem, animals cannot know what S+ is and must risk error by guessing. Eventually, they acquire a learning set (win-stay; lose-shift) that minimizes errors. Given an uncertainty response that produced a hint, animals might learn to respond uncertain on first trials and thus proceed errorlessly. This learning set – Uncertain? Ask! – would apply generally across stimulus contexts.

R5.2. Ecological uncertainty

Shettleworth & Sutton encouraged research on animals' spontaneous metacognitive activities in biologically relevant contexts. This idea was echoed by **Wilkins et al.**, who noted that one could explore animals' monitoring of dominance relations, alliances, face perception, and so forth. **Griffin** contributed lovely examples of the perceptual uncertainties that animals face. The possibility of ecological research is exciting. Imagine releasing a seed-caching bird into the laboratory environment to find the seeds it has hidden. Could it then use an uncertainty response to report that it is out of seeds it knows how to find? This could suggest the level of awareness the animal has about its seed memory and the level of control it exerts over the search process or its termination.

R5.3. Category uncertainty

Shimp asked whether animals could use a monitoring response to decline trials in which ambiguous stimuli require a category decision. This question intrigues us because of our interest in categorization (Smith 2002; Smith & Minda 1998; 2001). Shimp focused on the categorization paradigm used by Ashby and his colleagues (Ashby & Maddox 1998) and by Shimp (Herbranson et al. 1999) in his avian research.

R5.4. Varieties of cognitive monitoring

Shettleworth & Sutton wondered whether perceptual monitoring and memory monitoring necessarily go hand in hand. **Griffin** echoed this point by suggesting that simple perceptual consciousness is a core function of nervous systems. This capacity might therefore be more basic and more dispersed across species than memory monitoring. Dissociations of this kind would be illuminating if they emerged from precise cross-species comparisons.

R5.5. Information-seeking

Call described an information-seeking procedure in which animals behave adaptively by inspecting tubes to find the one containing food. This research recalls other research showing that dolphins echo-locate more intently and monkeys gaze longer when identifying difficult stimuli (Roitblat et al. 1990; Schrier & Wing 1973). We believe that information-seeking paradigms like Call's can be a constructive part of the overall picture of self-regulation by animals. However, it is still not known how high-level a cognitive capacity tube-inspection is.

R5.6. Other uncertainty

Call, **Flavell**, and **King** thought that research could exploit connections between uncertainty-monitoring and theory-of-mind (ToM) paradigms. Flavell asked what animals know about others' uncertainty states/reactions, whether they can monitor another's uncertainty, and whether they can use the knowledge that a conspecific knows something. Call outlined an other-uncertainty experiment. King pointed out that the guesser-knower ToM tasks are related to uncertainty-monitoring tasks. Flavell is correct that we would value research connecting these areas.

R5.7. Dissociating first- and second-order states

For **Browne**, the problem of metacognition research is that first-order states regarding stimuli often correlate with second-order states like uncertainty. The challenge is to dissociate the two and show that the second-order state affects behavior independently. It is worth thinking how this might be done. Could one find occasions when organisms are confident, even though the first-order stimuli (and associated reinforcement histories) indicate they should not be, or when they feared an error though they shouldn't? Would these results support the idea that second-order states help determine behavior?

R5.8. Confidence ratings by animals

Son et al. considered the interesting possibility that animals might make a metacognitive judgment after a perceptual response. Procedures for studying retrospective confidence judgments have now been developed by Son and Kornell (in press) and Shields et al. (submitted). Both procedures use a betting procedure to let participants accept higher reward/penalty stakes for the trials they strongly feel they know. Figure R3 illustrates some results. Humans favored high- and low-confidence responses, respectively, for easy trials near the ends of the trial continuum and for difficult trial near the middle (Fig. R3A). Humans showed better performance when they expressed high confidence than

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

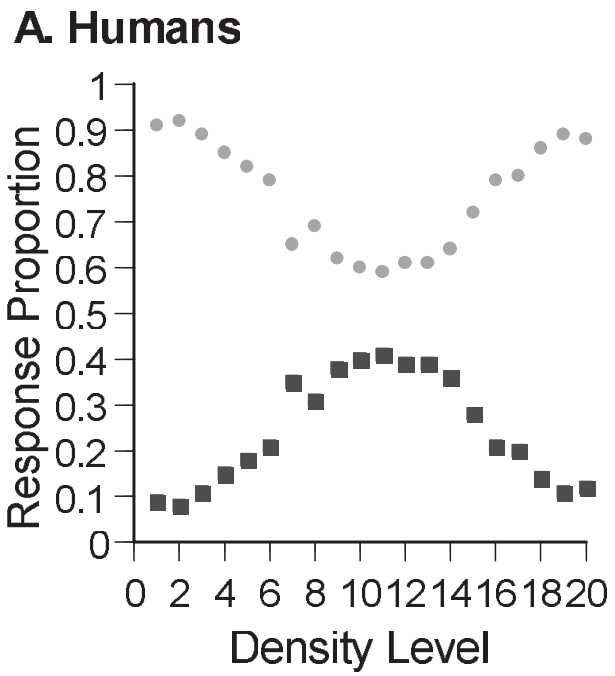


Figure R3A. The performance of humans in the confidence-rating task of Shields et al. (submitted). Humans first made a dense-sparse response without any feedback, then they were given two response options that let them select different reward/penalty stakes depending on their confidence in the answer they had given. In these four panels the 61-step density continuum has been binned three steps at a time. Density levels 1 to 10 (density steps 1 to 30) deserved the sparse response. Density levels 11 to 20 (density steps 32 to 61) deserved the dense response. The filled and open circles show humans' use of the high- and low-confidence responses, respectively, for stimuli at different density levels.

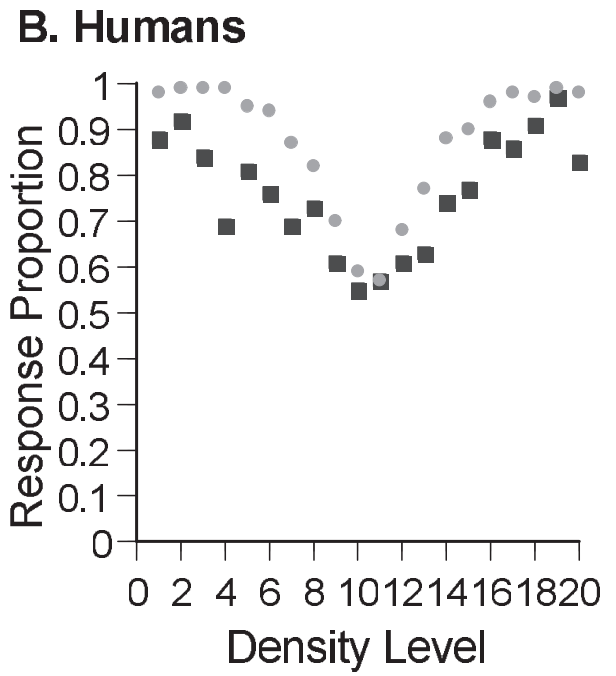


Figure R3B. Humans' proportion of correct responses at each density level when they chose the high- and low-confidence response.

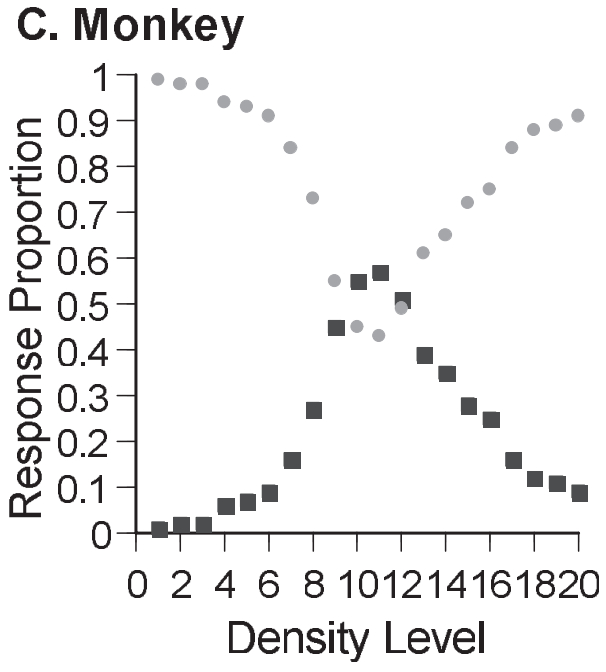


Figure R3C. The performance of a rhesus monkey in the confidence-rating task.

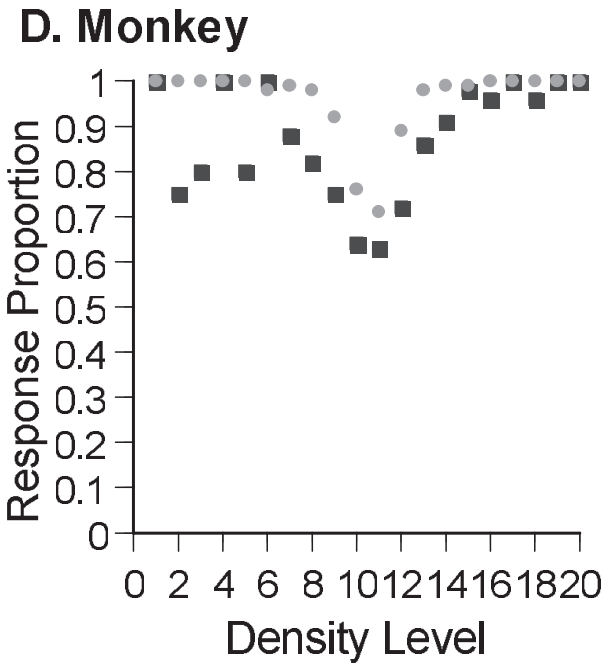


Figure R3D. The monkey's proportion of correct responses at each density level when it chose the high- or low-confidence response.

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

when they expressed low confidence – even with objective trial level controlled (Fig. R3B). Their feelings of confidence were veridical. Crucially, a monkey performed almost identically (Figs. R3C, R3D).

Metcalfe acknowledged that animals in the target article's experiments used a graded confidence scale of recognition but would not call this metacognition. Yet she seemed to accept as metacognitive graded judgments that were made after the cognitive response. This reflects her belief that the sequencing of behaviors determines what is metacognitive. In contrast, we believe that the underlying psychological processes are determinative, and section R3 discussed why the sequencing might be unimportant psychologically. In any case, the work of Son et al. and Shields et al. together seem to meet Metcalfe's standard and show that primates can make retrospective/metacognitive reports of uncertainty like those of humans.

R5.9. Pushing cognitively unsophisticated species harder

Shimp took it as a bitter pill that we were skeptical of pigeons' capacity for metacognition. His research on "metabehavior" could justify this caution because it might show a step toward metacognition (Shimp [1982; 1983] demonstrated that pigeons can respond discriminatively based on their recent behavior). **Zentall** was also unsettled on the species point and pointed to rats' trial-and-error behaviors to illustrate his point. Zentall also correctly pointed out that accepting the null hypothesis is uncomfortable scientifically. We reiterate (as in sects. 12, 12.1, and 12.2 of the target article) our agreement with the null-hypothesis point, our sense that the metacognitive capacity is not all-or-none, and our belief that methods in this area must be crafted to elicit the most sophisticated performances from animals they can achieve. This will help when animals are near their limits and might fall back to associative behavioral solutions. It will also give null results the best chance to be informative. For example, the range of forgetting intervals used by Inman and Shettleworth (1999) produced only a small, insignificant increase in safe-key use (sect. R2). Perhaps the motivation of the birds to escape trials or be metacognitive was not sufficient and this produced the null result.

R6. Considerations on formal modeling

Nelson suggested that our formal perspective used too strong a version of Signal Detection Theory (SDT). We assumed that the distributions (e.g., of trace strength) were equally variable, including the noise and the signal-plus-noise distributions. We agree this assumption is approximate and could make our measures of d′ approximate. The model we used was intuitive and accessible to the wide range of *BBS* readers. It accomplished a lot of data reduction with few parameters. Moreover, it did not seem that different variabilities between the distributions would change animals' psychological problem within a memory-monitoring task or the psychological interpretation of performance. This interpretation comes from the primary data pattern, and Nelson agreed that the underlying phenomena are best explained using metacognitive mechanisms.

Nevertheless, **Nelson** is correct that it will be valuable to gain more precision in modeling animals' representational spaces in these tasks. As he suggested, one can find the slope of the ROC curve by manipulating the payoff matrix and thus estimate the relevant variabilities. **Mazzoni** also considered manipulating the payoff matrix. Alternatively, one can plot the ROC curve by treating the three responses in an uncertainty-monitoring task as a rating scale of confidence (MacMillan & Creelman 1991, pp. 58–67). In one case, we did provide detailed SDT analyses, including variabilities and ROC slopes (see the Appendix in Smith et al. 1998). Nelson is correct that the slopes are less than 1 for humans and monkeys. Perhaps this convergence strengthens the isomorphism between the memory-monitoring performances of the two species.

Goldsmith & Koriat seemed comfortable with the idea that animals have a rudimentary capacity for metacognition and that they withhold responses when they feel uncertain and thereby increase the accuracy of offered memory reports. However, they argued that the distinction between monitoring and control was not addressed sufficiently by our SDT model or by research on animal metacognition. We agree with this comment, too.

This is our understanding of the problem. Presumably a memory item will have a level of activation in the animal's mind that could recommend a *not there*, *uncertain*, or *there* response. The animal chooses one of these. But why? There are two intervening steps. The animal has to *monitor* the trace strength and then let that trace strength *control* behavior. The criterion lines in our SDT model could reflect either the monitoring or control process.

Addressing this concern with humans, the commentators obtain forced-report and free-report responses to different memory items, and a confidence rating for each forced-report answer (see the Quantity-Accuracy-Profile assessment in Koriat & Goldsmith 1996a; 1996b). Eventually it might become possible to assess the components of metacognitive performance for animals, too. This assessment will benefit from the research on animals' confidence ratings that we discussed in section R5.8. (We believe the confidence-rating approach may be more robust and comparable across species than relying on species-natural uncertainty behaviors.) This assessment will also benefit from the development of flexible, animal-friendly, free-report measures of memory (e.g., Menzel 1999; Schwartz 2002).

Son et al. also made the monitoring-control distinction and suggested that it had not been sufficiently addressed regarding animals. They noted that animals sometimes control behavior poorly even when they have good knowledge. They suggested that different species might have strengths in different aspects of metacognition. This recommends the broad comparative approach that we also endorse. It also recommends finer distinctions within the metacognitive capacity (e.g., monitoring vs. control) as suggested by **Goldsmith & Koriat.** Such distinctions might also refine the construct of metacognition and define the boundaries of metacognitive abilities.

Metcalfe and **Shettleworth & Sutton** made a serious error regarding our model. We did not say that the usefulness of signal detection theory or the fit of signal detection models supported the metacognitive interpretation. We said the model

> offers a neutral description of performance that is inclusive theoretically because it makes no theoretical commitments toward behaviorism or cognitivism. It clarifies the formal structure of behavior so that different theoretical perspectives can be brought to bear on it. (target article, sect. 13)

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

The interpretation of behavior comes from the primary data pattern. Most commentators endorsed the metacognitive interpretation while raising interesting questions about consciousness and self-awareness. (A small point is that Shettleworth & Sutton did not summarize p. 52 of MacMillan and Creelman [1991] to our satisfaction. That passage says that humans can change criteria consciously. This is why instructions work! The only point was that SDT – a mathematical formulation – "takes no stand" on consciousness. Of course not – consciousness is a psychological question. For the same reason, it is not constructive to ask, as Metcalfe did, whether introductory SDT descriptions attribute the underlying processes to metacogniton. Of course they don't – metacognition is a psychological question.)

It will facilitate understanding the next sections of the response if we summarize the target article using these nine points:

1. Animals perform nearly identically to humans in the experiments described in the target article, producing some of the strongest existing human-animal performance similarities.
2. Some of the performances cannot be explained by low-level, associative mechanisms:
3. Instead, they require more controlled and decisional cognitive processing.
4. Humans attribute their uncertainty responses to conscious metacognition.
5. The parsimonious interpretation is that the performances by humans and animals have a common psychological basis:
6. Thus, animals show the functional analog to conscious metacognition in humans.
7. However, animals' performance needn't imply their full declarative consciousness.
8. Their performance also needn't imply their full self-awareness.
9. Still, there is probably an adaptive advantage favoring animals who can – in the immediate moment of difficulty and uncertainty – construct a decisional assemblage that lets them find creative, emergent solutions to novel cognitive problems. A working consciousness would serve this function well.

The commentators focused on a number of these points.

R7. Associative interpretations do not suffice

Several commentators (**Shettleworth & Sutton**, **Wilkins et al.**, **Zentall**) followed animal psychology's historical tendency to demote animals' performance. They re-asserted that animals' uncertainty responses could be associatively learned or responsive to stimuli. Wilkins et al. said that animals could be learning a third response associated with shorter timeouts. Zentall said that animals use a primary or uncertainty response, respectively, as discriminability is high or low. Shettleworth & Sutton suggested that animals learn the relative reward rates for opting out versus taking the test in the presence of external stimuli.

These incorrect interpretations prompt four points in reply. First, the target article (sect. 14.1) made clear why the stimulus-based interpretation cannot explain the most important results involving same-different judgments and memory monitoring. Second, although the target article acknowledged that a couple of tasks did allow an associative interpretation, it serves the literature poorly to focus on this to push the associative account. This misuses the effort the target article took to reason carefully. (Zentall also claimed that only one monkey showed the metacognitive phenomenon in Smith et al. [1997]. In fact both did in that article's main task, and one did in the subsidiary task.)

Third, given tasks that require a high-level cognitive interpretation and others that do not, it is not necessarily good science to emphasize the unparsimonious account that explains the tasks in qualitatively different ways. If we already know that animals know when they don't remember, why shouldn't they also know when they can't tell dense from sparse? Fourth, we caution that sometimes interpretations sound low level and behaviorist but are not. For example, **Zentall** suggested that animals choose uncertain responses when the discrimination is difficult. But who decides this? The explanation confirms that a monitoring utility is at work.

Browne also favored a low-level interpretation of the target performances. He stated that a subject who is uncertain whether two tones are the same or different is in a first-order state. Unfortunately, his idea of first-order went undefined. Given one possible meaning, we disagree. Uncertainty about a relation between two stimuli is cognitively a third-order state because the uncertainty (3) is about a relation (2) between perceptual representations (1). Browne grants that the uncertainty would be higher order if the subject is uncertain whether the experiences of hearing the two tones are different, but not if the subject is uncertain whether the two tones are different. But organisms can never compare tones – they are Kant's things-in-themselves. Because subjects can only compare perceptual experiences, one concludes again that uncertainty about the relation between two tones is higher order.

R8. Metacognition without self-awareness

Carruthers sought a middle ground between associative control by stimulus cues and metacognitive control by uncertainty cues. (**Call**'s interpretation also fell between low-level behavioristic and high-level metarepresentational explanations. One can read **Metcalfe**'s acknowledgment that animals use a graded "confidence" scale as a middle-ground interpretation, too.) Carruthers granted that the target performances were not associative but genuinely cognitive. He granted that animals have degrees of belief that tones are high or low and boxes dense or sparse, and moreover have ways of telling when the state of belief is insufficient to warrant a primary discrimination so that the uncertainty response can be chosen instead. This sounds metacognitive. However, Carruthers's objection to this characterization was that the animal needn't "believe *of* itself that it has a desire with a certain strength," and needn't "represent itself as lacking a sufficient degree of belief."

Proust agreed that animals' hierarchical cognitive control need not be self-reflexive (in the sense of using an integrated representation of one's own mental, social, and physical dispositions). We hope the target article nowhere claimed that animals have an awareness of the self's being uncertain, or that they know they are minds who are uncertain. We neither claim that they do nor claim that they don't.

R9. Metacognition without consciousness

Son et al. and **Mazzoni** endorsed that animals are showing a metacognitive performance in the target tasks. Maz-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

zoni found the demonstration of metacognition convincing. Son et al. acknowledged the demonstrations that animals can inform us that they feel uncertain, one proposed measure of metacognitive ability. However, they suggested that metacognition could be less declarative and conscious and more implicit. Both chose similar examples of implicit metacognition (**Hampton** did so, too). Son et al. noted that participants sometimes say they will know something before knowing the answer. Mazzoni noted that participants sometimes make metacognitive judgments using experience-based heuristics (e.g., "I know art history, so I'll know this"). A small point is that these domain-based metacognitive judgments do not apply to the performances described in the target article because these occur in only one domain at a time. Thus, animals cannot use a domain heuristic but must really evaluate knowing on a trial-by-trial basis. It would be interesting to run a several-domain task with animals and ask whether they make domain-based judgments of knowing.

Proust also took the implicit tack. He granted that animals show hierarchical control between processes, reflexivity at the task level, and metacognition. However, he cast narrowly what metacognition means, reducing it to something that is merely executive, architecture-bound, or structural. He suggested that this procedural metacognition might only represent the system's current epistemic states but not the system's attitudinal contents. And if so, metacognition would not be semantic, intentional, or metarepresentational.

Flavell synthesized the issues of consciousness and self-awareness into his typology of uncertainty monitoring. First, organisms might have a psychological uncertainty reaction that does not contain a conscious feeling of uncertainty or a conscious feeling of any kind. Second, there might be a conscious feeling of uncertainty or hesitancy, but the creature would not reflect on it or identify it as being that kind of feeling. Third, there might be uncertainty feelings and conscious awareness that it is an uncertainty feeling – that is, the creature would feel unsure and know that the feeling it is experiencing is that of being unsure. Flavell hypothesized that only the first and second of these occur in animals and human infants, whereas all three occur in older humans

We hope the target article seemed cautious (e.g., sect. 15) regarding consciousness and that we did not assume animals' conscious metacognition as it seemed to **Mazzoni** we did. Nonetheless, for the record, we definitely acknowledge the phenomenon of implicit metacognition, and we agree that an uncertainty-monitoring performance might occur without full consciousness. This important issue is discussed further in section R12.

R10. Metacognition and the "epiphenomenon" problem

One seeming human–animal difference is that we can ask humans afterwards, and they report uncertainty and consciousness. It is a common idea that humans' metacognition is more real and higher level because it is so declarative.

This idea is incorrect for two reasons. First, this standard depends on humans' making a declaration that animals cannot. A soccer referee denied a yellow card would still see fouls but could not declare them. Similarly, animals might be consciously metacognitive in our tasks but not say so.

Second, this standard depends on secondary reports by humans. Yet five commentators suggested that these reports are not so probative. **Wilkins et al.** noted that humans might attribute undeserved meaning to uncertainty responses given social desirability and incomplete understanding of their mental processes. **Browne** noted that humans go reflective and metacognitive when probed but that the cognitive state that drove the behavior might not have been so. **Carruthers** and **Metcalfe** agreed that the fact that humans' explanations of their feelings and behaviors are metacognitive doesn't guarantee that the decision-making process itself was metacognitive. (Metcalfe made a mistake, though, by claiming that there was no special metacognitive priority given to the uncertainty response. The uncertainty verbalizations were qualitatively different and uniquely metacognitive – see sections 5 and 6 in the target article.) **Zentall** thought that metacognition might not precede a difficult choice but follow it to help justify it and reduce the aversion associated with an imperfect outcome.

These comments are constructive for several reasons. First, becoming clear on the status of secondary reports could help define the construct of human metacognition. Second, these comments help in interpreting animals' metacognitive performances. **Marino** worried that we might hold animals to a higher consciousness standard than we do humans or try to place their performance in a theoretical space that does not exist. The problem with secondary reports explains Marino's worry. We could withhold consciousness from animals unless they can say so when we ask. But if these introspections are not probative anyway, then demanding them from animals is inappropriate and involves a theoretical space that is artificial.

Third, by dismissing secondary reports, these commentators reduced the species differences in performance. Shorn of the introspections, there is wide agreement across the commentaries that the underlying processes have a common psychological basis. Fourth, perhaps the shearing is a good thing, because the problem of the secondary reports is pretty self-defeating. One can't fix the problem just by asking for introspections on every trial or even within a trial. The real point is that probing – period – changes the phenomenon, Heisenberg style.

Fifth, therefore, these comments move one away from secondary reports that artificially create species distance, that may change the observed phenomena, and that may involve metaconsciousness, not metacognition. Instead, they put the focus back on the character of information processing when humans or animals face uncertain situations. In section R12 we discuss why this information processing deserves close cognitive-comparative scrutiny for reasons apart from secondary introspections.

R11. Metacognition and other executive, self-regulatory functions

Some commentators found connections between uncertainty-monitoring/metacognition and cognitive self-regulation/executive function. We value these connections and the commentators' ideas about them. For example, **Rumbaugh et al.** discussed the problem animals face when they need new responses with no antecedent in previously learned behaviors. The behavioral solutions that animals contrive under these conditions are called emergents. By a

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

logic we endorse, they linked uncertainty behaviors and emergents. These would occur within the same uncertain cognitive situations. The uncertainty would slow the animal down, granting time for cognition that produced emergents. The uncertainty might also move information processing to a more controlled or even conscious level where the emergent would be constructed. As Rumbaugh et al. note, the historically important emergent called insightful behavior has always been observed to be preceded by behavioral hesitation.

Fantino considered the areas of self-control and information seeking. Pigeons are impulsive and delay gratification poorly. But pigeons (and humans) will sometimes make a "save me from my impulsiveness" response by which they commit to wait for the bigger reward. Pigeons (and humans) will also make observing responses (which have no reinforcement implication) that tell them what reinforcement schedule is in effect. They seem to like to know. Fantino finally asked about the relationship between metacognition and self-control/information seeking. Do commitment responses show that the animal knows its impulsiveness and how to guard against it? Do observing responses show a metacognitive search for information? Depending on the answers, one does see phylogenetic similarities between pigeons and humans in areas related to metacognition, just as one does with the meta-behavior experiments of **Shimp.**

R12. The evolution of metacognition

Proust asked how explicit, conscious metacognition might have evolved. She reasoned that animals began with an implicit control system that transitioned to consciousness when the animal gained declarative access to the control system that had been in the cognitive unconscious (Karmiloff-Smith 1992; Rozin 1976).

But why did this transition occur? What fitness advantage did conscious metacognition confer? This is a fair question. **Shimp** noted that many behaviors are executed more fluently if they are done automatically and unconsciously. (One can be eaten if one consciously reflects on the wind sound through an owl's wings.) **Rumbaugh et al.** agreed that animals can rely on well-learned behaviors when the world is certain. **Proust** added that implicit forms of cognitive control are more economical. The idea of implicit metacognition raises the same point. If it is so good and useful, why would we have explicit metacognition?

The target article (sect. 15) tried to say why. Life often brings difficult, uncertain situations for which automatic habits do not suffice. **Griffin** made this point beautifully. Then, in the moment, the creature must decide what to do. These decisions will use higher-level cognition, including immediate memory to hold useful information and behavioral possibilities. We concluded that a working consciousness is the perfect cognitive utility for holding this decisional assemblage in immediate processing awareness. **King** added that uncertain situations could have an emotional color of threat or arousal. We agree this would provide an adaptive signal to the organism to go and think carefully. Gray (1995) described the neural circuits that may arrest behavior, increase arousal, and redirect attention and mental resources toward the causes of uncertainty and difficulty (see also Smith 1995).

Rumbaugh et al. agreed that organisms go to a different cognitive level when they encounter ambiguous and uncertain situations. This is when emergent behaviors occur (sect. R11). But these commentators would not commit to this level's being conscious. It is an interesting question whether an unconscious automaton could ever contrive an emergent behavioral solution. We wonder whether Rumbaugh et al. would say yes. **Griffin** agreed that the uncertainties of life that are critical for survival are the things that will most require conscious attention.

Marino asked directly whether the responses by animals that look metacognitive would work without declarative consciousness. She did commit. Her answer was that we provided substantial evidence and a cogent argument for declarative consciousness as a shared psychological property in humans, monkeys, and dolphins.

We will also commit. At issue is a distinctive cognitive utility wherein one holds in immediate memory the decisional assemblage to process a difficult situation and choose a first-time behavior or else seek help or escape. This processing occurs in the immediate mental moment. The situation might be novel so that no training comes to bear. Judgment, decision making, and cognitive control are required instead. If there is any moment in information processing that would use conscious processing, this is it. If this situation can be handled implicitly and unconsciously, there is no time when consciousness would be needed. Accordingly, we think it is essentially definitional that the cognitive place that holds this distinctive kind of cognitive processing would be called working consciousness.

Now, one can deny this cognitive utility another eye (metaconsciousness) that examines the process and declares or feels the states. One can deny this cognitive utility the feeling of being a self who is uncertain and in trouble. But to us, that utility would still have everything that immediate conscious processing is supposed to have, and it would still deserve the label we have given it.

The adaptiveness of working consciousness helps explain why we have explicit metacognition. And if it is adaptive for the human mind to have evolved explicit metacognition, then other minds probably have evolved it, too.

R13. The development of metacognition

As **Flavell** noted, questions remain about children's earliest metacognitive attainments. One problem is that children have often been given uncertainty-monitoring tasks that are difficult and verbal and that may underestimate their metacognitive capacities. Simple perceptual tasks might be better for revealing uncertainty monitoring in very young children. Thus we agree with Flavell's suggestion that the target tasks, built to suit animal participants, might also help find the earliest roots of children's metacognition. Here we mention one child we tested who used our star response to decline trials, and whom we asked to explain that response. He executed a painful, 30sec, 360-degree fidget while he struggled to say what he had just done fine. Flavell gave a similar example from the block-building study. These perceptual tasks of metacognition, in relation to others in the literature, do raise interesting questions about levels of consciousness in metacognitive performance and about which species and ages have which levels.

Campos & Karmiloff-Smith discussed other important

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

rationales for studying metacognition's developmental time course. First, understanding the phenotypic outcome of an ability may require understanding the ability's developmental trajectory. In turn, this can help one judge whether similar cross-species behaviors are underpinned by similar processes. The developmental perspective also lets one see whether metacognitive capacities interact over developmental time with other cognitive processes and whether this interaction is the same across species.

We agree with these points and offer one example from our perspective. An influential view is that children's metacognition (along with other self-regulatory functions) is handed off from the caregiver during the time in which he or she serves as the child's regulator and metacognitive agent. If metacognition is the internalized cognitive guidance of the parent, one will not see it in monkeys who don't parent in that way. Thus, one wonders whether some forms of metacognition have an endogenous, not dyadic, origin, so that they could arise in solitary species, in social but nonteaching species like monkeys, and in children independent of the parent's cognitive modeling.

R14. Parsimony in comparative studies of cognition

Zentall argued that the parsimonious explanation of performance was associative and reinforcement based. Section R7 discussed why this explanation fails. Moreover, Zentall's interpretation is doubly unparsimonious. He focuses on a subset of the relevant performances in pursuing the associative account but then requires a second mechanism to explain the most important data. He focuses only on the animals' performances but then requires a different mechanism to explain humans' metacognitive performances. His is not a true comparative perspective because it is too task- and species-narrow. Yet, the issue is truly comparative because it involves several animal species including humans.

In contrast, many commentators valued (within limits) a cross-species parsimony that granted humans' and monkeys' performances a common processing basis. **King** related this idea to de Waal's (1996) idea of "evolutionary parsimony." **Rumbaugh et al.** noted that animal and human uncertainty responses might be consistent psychologically. **Shimp** noted that the historical brand of parsimony can do mischief and could be counterproductive regarding the area of animal metacognition. **Campos & Karmiloff-Smith** agreed that normally one would take similar behavior between different groups or species as indicating similar processes. **Marino** endorsed the idea of an integrative parsimony that allowed for common processing mechanisms (especially given phylogenetically close species with homologous brain structures) on some levels and possible differences on other levels.

King, though, saw the lack of parsimony if one explained monkeys' and humans' performances in qualitatively different ways, but less so if one made the interpretation that monkeys have a lesser or precursor capacity for metacognition. We are comfortable with and interested in the idea of "junior" metacognition.

Campos & Karmiloff-Smith sounded the caution that sometimes apparently similar cognitive performances (e.g., by Williams Syndrome children) turn out, given in-depth studies, to reflect different underlying processes. We agree with the caution. Our point was just about the proper default assumption of science (that the same thing is the same thing) and the problem that behavioral psychologists sometimes automatically assume a qualitatively different explanation of animals' performance. We encourage the in-depth studies that could show what aspects of metacognition monkeys do and do not have.

The only critical thing to us is that humans and monkeys be placed on the same interpretative playing field, so that one compares them carefully even if one grades them with different levels of consciousness or selfhood. Doing so is true comparative psychology, and it can lead to rapid theoretical progress. We hope that researchers of all persuasions will look at all the commentaries and see that this dialog is almost the closest theoretical look that has ever been taken at a human-animal cross-species similarity in performance. With the discussion occurring on this level, we have no position or conclusion to defend. Our position is just that the discussion should have this level and this character, and it clearly does.

R15. Metacognition and theory of mind (ToM)

Several commentators (**Call**, **Flavell**, **King**, **Proust**, **Wilkins et al.**) discussed the possible relationships between metacognition and theory of mind (ToM). These capacities might be uncorrelated aspects of mind with no functional dependency or phylogenetic co-occurrence. Or, there might be a directional connection from metacognition to ToM, with self-awareness evolved first as the substrate for other-awareness. **King** raised this precursor-capacity possibility. Third, metacognition and ToM could be symptoms of the flowering of mind that produced consciousness, self-awareness, and metacognition just once in the human-ape evolutionary line. The three possibilities predict different comparative maps regarding metacognition and theory of mind. First, one could see a patternless crazy quilt of the capacities sprinkled across species. Second, one might see that all ToM species have the precursor metacognition, but the reverse would not hold. Third, one could see that apes and humans – only – have metacognition and ToM.

We have several reflections on this issue. First, this is an empirical question. One could fill in this comparative map through research. Second, this research depends on independent assays of metacognition and ToM that can be used across species and that have consensual support in the field. An important purpose of the target article was to move toward these consensual measures of metacognition. Third, the need for independent assays of the two capacities is important even if one prefers Gallup's idea that metacognition and ToM emerged together within the anthropoid mind (cf. Gallup 1970; 1982). The more intriguing and influential this idea, the more it deserves the strong test allowed by separate measures of separate capacities.

Fourth, some commentaries almost seemed to let the Gallup perspective transcend empirical test. **Proust** argued that because monkeys lack a ToM and fail the dye-mark test, it is highly implausible that they could be metacognitive. **Wilkins et al.** argued that, given monkeys' dye-mark and tactical-deception failures, they should not even have been our featured participants in metacognition experiments. They further argued that monkeys' poor use of mir-

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

ror-guided hand behaviors to gain hidden food rewards was predictive of their lack of metacognition. They further stated that the model and methods we used were appropriate and interesting but only when applied to humans. We don't understand why, and we find this statement problematic.

We caution that the perspective behind these comments could impede empirical and theoretical progress in the areas of awareness, metacognition, and theory of mind. The perspective conflates these capacities because it assumes they all reflect the same evolutionary flashpoint. It discourages research on differentiated areas of awareness. It slows the development of separate assays for separate capacities because a measure of one should be a measure of all. It views skeptically a close analysis of the cognitive content of the intriguing dye-mark test. It prejudges the outcome of metacognition experiments with monkeys. It argues that monkeys should not even be featured in metacognition experiments. It undermines the goal of tracing the phylogenetic co-occurrences and the functional dependencies between metacognition and theory of mind because the comparative map is theoretically predrawn. We urge that the perspective expressed by **Wilkins et al.** be treated as an interesting scientific possibility but not as an article of faith that bends inquiry. In fact, given the evidence for metacognition in monkeys and dolphins, and the demonstration that dolphins pass the mirror-dye test (Reiss & Marino 2001b), the comparative map may not finally take the form predicted from this perspective.

ACKNOWLEDGMENT

The preparation of this article was supported by the National Institutes of Child Health and Human Development (HD-38051).

References

Letters "a" and "r" appearing before authors' initials refer to target article and response respectively.

Ainslie, G. W. (1974) Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior* 21:485–89. [EF]

Alsop, B. (1998) Receiver operating characteristics from nonhuman animals: Some implications and directions for research with humans. *Psychonomic Bulletin and Review* 5:239–52. [TON]

Amsterdam, B. (1972) Mirror image reaction before the age of two. *Developmental Psychobiology* 5:297–305. [VMW]

Anderson, J. R. (1994) The monkey in the mirror: The strange conspecific. In: *Self-awareness in animals and humans: Developmental perspectives,* ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [VMW]

Anderson, J. R. & Bower, G. H. (1973) *Human associative memory.* Wiley. [CPS]

Anderson, J. R. & Gallup, G. G., Jr. (1997) Self-recognition in *Sanguinus:* A critical essay. *Animal Behaviour* 54:1563–67. [JP]

Angell, F. (1907) On judgments of "like" in discrimination experiments. *American Journal of Psychology* 18:253. [aJDS]

Ashby, F. G. & Maddox, W. T. (1998) Stimulus categorization. In: *Measurement, judgment, and decision making: Handbook of perception and cognition,* ed. M. H. Birnbaum. Academic Press. [CPS, rJDS]

Atkinson, R. C. & Juola, J. F. (1974) Search and decision processes in recognition memory. In: *Learning, memory, and thinking,* ed. D. H. Krantz, R. C. Atkinson, R. D. Luce & P. Suppes. Freeman. [aJDS]

Au, W. W. & Moore, P. W. (1990) Critical ratio and critical bandwidth for the Atlantic bottlenose dolphin. *Journal of the Acoustical Society of America* 88:1635–38. [aJDS]

Banks, W. P. (1970) Signal detection theory and human memory. *Psychological Bulletin* 74:81–99. [MG]

Bargh, J. A. (1997) The automaticity of everyday life. In: *Advances in social cognition, vol. 10,* ed. R. S. Wyer, Jr. Erlbaum. [JP]

Baron-Cohen, S., Ring, H., Moriarty, J., Schmitz, D., Costa, D. & Ell, P. (1994) Recognition of mental state terms: A clinical study of autism and a functional neuroimaging study of normal adults. *British Journal of Psychiatry* 165:640–49. [VMW]

Bekoff, M., Allen, C. & Burghardt, G. M., eds. (2002) *The cognitive animal: Empirical and theoretical perspectives on animal cognition.* MIT Press. [DRG, GM]

Bitterman, M. E. (1960) Toward a comparative psychology of learning. *American Psychologist* 15:704–12. [TRZ]

Blough, D. S. (1958) A method for obtaining psychophysical threshold from the pigeon. *Journal of the Experimental Analysis of Behavior* 1:31–43. [aJDS]

Bolles, R. & Petrinovich, L. (1954) A technique for obtaining rapid drive discrimination in rats. *Journal of Comparative and Physiological Psychology* 47:378–80. [JEK]

Boneau, C. A. & Cole, J. L. (1967) Decision theory, the pigeon, and the psychophysical function. *Psychological Review* 74:123–35. [aJDS]

Boring, E. G. (1920) The control of attitude in psychophysical experiments. *Psychological Review* 27:440–52. [aJDS]

Boysen, S. T., Berntson, G. G., Hannan, M. B. & Cacioppo, J. T. (1996) Quantity-based interference and symbolic representations in chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes* 22:76–86. [LKS]

Brown, A. L., Bransford, J. D., Ferrara, R. A. & Campione, J. C. (1982) Learning, remembering, and understanding. In: *Handbook of child psychology, vol. 3,* ed. J. H. Flavell & E. M. Markman. Wiley. [aJDS]

Brown, A. S. (1991) A review of the tip-of-the-tongue experience. *Psychological Bulletin* 109:204–23. [aJDS]

Brown, W. (1910) The judgment of difference. *University of California Publications in Psychology* 1:1–71. [aJDS]

Burke, D., Cieplucha, C., Cass, J., Russell, F. & Fry, G. (2002) Win-shift and win-stay learning in the short-beaked echidna (*Tachyglossus aculeatus*). *Animal Cognition* 5:79–84. [LKS]

Byrne, R. W. & Whiten, A. (1991) Computation and mindreading in primate tactical deception. In: *Natural theories of mind: Evolution, development and simulation of everyday mindreading,* ed. A. Whiten. Basil Blackwell. [aJDS]

Call, J. (2001) Chimpanzee social cognition. *Trends in Cognitive Sciences* 5:369–405. [JC]

(in press) The self and the other: The missing link in comparative social cognition. In: *The evolution of consciousness in animals and humans,* ed. H. Terrace & J. Metcalfe. Oxford University Press. [JC]

Call, J. & Carpenter, M. (2001) Do chimpanzees and children know what they have seen? *Animal Cognition* 4:207–20. [JC, SJS]

Call, J. & Tomasello, M. (in press a) Social cognition. In: *Primate psychology,* ed. D. Maestripieri. Harvard University Press. [JC]

(in press b) What chimpanzees know about seeing revisited: An explanation of the third kind. In: *Joint attention,* ed. N. Eilan, C. Hoerl, T. McCormack & J. Roessler. Oxford University Press. [JC]

Carter, D. E. & Werner, T. J. (1978) Complex learning and information processing by pigeons: A critical analysis. *Journal of the Experimental Analysis of Behavior* 29:565–601. [aJDS]

Case, D. A., Fantino, E. & Wixted, J. (1985) Human observing: Maintained by negative informative stimuli only if correlated with improvement in response efficiency. *Journal of the Experimental Analysis of Behavior* 43:289–300. [EF]

Castro, C. A. & Larsen, T. (1992) Primacy and recency effects in nonhuman primates. *Journal of Experimental Psychology: Animal Behavior Processes* 18:335–40. [aJDS]

Cheney, D. L. & Seyfarth, R. M. (1990) *How monkeys see the world.* University of Chicago Press. [JP, aJDS]

Clayton, N. S., Griffiths, D. P. & Dickinson, A. (2000) Declarative and episodic-like memory in animals: Personal musings of a Scrub Jay. In: *The evolution of cognition,* ed. C. Heyes & L. Huber. MIT Press. [CPS]

Clayton, N. S., Griffiths, D. P., Emery, N. J. & Dickinson, A. (2001) Elements of episodic-like memory in animals. *Philosophical Transactions of the Royal Society, London B* 356:1483–91. [SJS]

Cohen, L. J. (1993) *An essay on belief and acceptance.* Oxford University Press. [PC]

Colombo, M. & Frost, N. (2001) Representation of serial order in humans: A comparison to the findings with monkeys. *Psychonomic Bulletin and Review* 8:262–69. [GM]

Commons, M. L., Nevin, J. A. & Davison, M. C., eds. (1991) *Signal detection: Mechanisms, models, and applications.* Erlbaum. [aJDS,SJS]

Costell, A. (1993) How Lloyd Morgan's *Canon* backfired. *Journal of the History of the Behavioral Sciences* 29:113–22. [JEK]

Cowey, A. & Stoerig, P. (1992) Reflections on blindsight. In: *The neuropsychology of consciousness,* ed. D. A. Milner & M. D. Rugg. Academic Press. [aJDS]

(1995) Blindsight in monkeys. *Nature* 373:247–49. [aJDS]

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

Cumming, W. W. & Berryman, R. (1961) Some data on matching behavior in the pigeon. *Journal of the Experimental Analysis of Behavior* 4:281–84. [aJDS]
Danion, J. M., Gokalsing, E., Robert, P., Massin-Krauss, M. & Bacon, E. (2001) Defective relationship between subjective experience and behavior in schizophrenia. *American Journal of Psychiatry* 158:2064–66. [MG]
Davison, M., McCarthy, D. & Jensen, C. (1985) Component probability and component reinforcer rate as biasers of free-operant detection. *Journal of the Experimental Analysis of Behavior* 44:103–20. [aJDS]
Dennett, D. (1979) How to change your mind. In: *Brainstorms,* ed. D. Dennett. Harvester Press. [PC]
Deruelle, C., Mancini, J., Livet, M. O., Casse-Perrot, C. & de Schonen, S. (1999) Configural and local processing of faces in children with Williams syndrome. *Brain and Cognition* 41:276–98. [RC]
De Waal, F. B. M. (1991) The chimpanzee's sense of social regularity and its relation to the human sense of justice. *American Behavioral Scientist* 34:335–49. [JEK]
(1996) *Good natured: The origins of right and wrong in humans and other animals.* Harvard University Press. [JEK]
De Waal, F. B. M. & Luttrell, L. M. (1988) Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology* 9:101–18. [JEK]
Dewey, J. (1934/1980) *Art as experience.* Perigee Books. [aJDS]
Dickinson, A. & Balleine, B. W. (2000) Causal cognition and goal-directed action. In: *The evolution of cognition,* ed. C. Heyes & L. Huber. MIT Press. [DRG]
Dinsmoor, J. A. (1983) Observing and conditioned reinforcement. *Behavioral and Brain Sciences* 6:693–728. [EF]
Donnai, D. & Karmiloff-Smith, A. (2000) Williams syndrome: From genotype through to the cognitive phenotype. *American Journal of Medical Genetics: Seminars in Medical Genetics* 97(2):164–71. [RC]
Dunlosky, J. & Nelson, T. O. (1992) Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory and Cognition* 20:374–80. [arJDS]
(1997) Similarity between the cue for judgments of learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language* 36:34–39. [aJDS]
Elliott, R. & Dolan, R. J. (1999) Differential neural responses during performance of matching and nonmatching to sample tasks at two delay intervals. *The Journal of Neuroscience* 19:5066–73. [aJDS]
Epstein, R. (1996) *Cognition, creativity, and behavior: Selected essays.* Greenwood. [EF]
Etkin, M. & D'Amato, M. R. (1969) Delayed matching-to-sample and short-term memory in the capuchin monkey. *Journal of Comparative and Physiological Psychology* 69:544–49. [aJDS]
Evans, J. & Over, D. (1996) *Rationality and reasoning.* Psychology Press. [PC]
Fantino, E. (1966) Immediate reward followed by extinction vs. later reward without extinction. *Psychonomic Science* 6:233–34. [EF]
(1977) Conditioned reinforcement: Choice and information. In: *Handbook of operant behavior,* ed. W. K. Honig & J. E. R. Staddon. Prentice Hall. [EF]
Fantino, E. & Case, D. A. (1983) Human observing: Maintained by stimuli correlated with reinforcement but not extinction. *Journal of the Experimental Analysis of Behavior* 40:193–210. [EF]
Farthing, G. W. & Opuda, M. J. (1974) Transfer of matching-to-sample in pigeons. *Journal of the Experimental Analysis of Behavior* 21:199–213. [aJDS]
Fernberger, S. W. (1914) The effect of the attitude of the subject upon the measure of sensitivity. *American Journal of Psychology* 25:538–43. [aJDS]
(1930) The use of equality judgments in psychophysical procedures. *Psychological Review* 37:107–12. [aJDS]
Flavell, J. H. (1979) Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist* 34:906–11. [aJDS,RRH]
Flavell, J. H., Speer, J. R., Green, F. L. & August, D. L. (1981) The development of comprehension monitoring and knowledge about communication. *Monographs of the Society for Research in Child Development* 46, Serial No.192. [JHF]
Fobes, J. L. & King, J. E. (1982) *Primate behavior.* Academic Press. [aJDS]
Frankish, K. (1998) A matter of opinion. *Philosophical Psychology* 11:423–42. [PC]
Frederick, D. A. & Platek, S. M. (under review) The chimpanzee self-awareness and theory of mind debate and the neurological underpinnings of self. *Trends in Cognitive Sciences.* [VMW]
Fujita, K. (1982) An analysis of stimulus control in two-color matching-to-sample behaviors of Japanese monkeys (*Macaca fuscata*). *Japanese Psychological Research* 24:124–35. [aJDS]
Gallup, G. G., Jr. (1970) Chimpanzees: Self-recognition. *Science* 167:86–87. [rJDS, VMW]
(1982) Self-awareness and the emergence of mind in primates. *American Journal of Primatology* 2:237–48. [arJDS,VMW]
Gallup, G. G., Jr. & Anderson, J. R. & Platek, S. M. (in press) Self-awareness, social intelligence, and schizophrenia. In: *The self and schizophrenia: A neuropsychological perspective,* ed. A. S. David & T. Kircher. Cambridge University Press. [VMW]
Gallup, G. G., Jr., Anderson, J. & Shillito, D. (2000) The mirror test. In: *The cognitive animal,* ed. M. Bekoff, C. Allen & G. Burghardt. MIT Press. [VMW]
Gallup, G. G., Jr. & Suarez, S. D. (1986) Self-awareness and the emergence of mind in humans and other primates. In: *Psychological perspectives on the self, vol. 3,* ed. J. Suls & A. Greenwald. Erlbaum. [aJDS]
Gazzaniga, M. (1998) *The mind's past.* California University Press. [PC]
George, S. S. (1917) Attitude in relation to the psychophysical judgment. *American Journal of Psychology* 28:1–38. [aJDS]
Gibson, E. J. & Walk, R. D. (1960) The "visual cliff." *Scientific American* 202:64–71. [JHF]
Gilden, D. L. & Wilson, S. G. (1995) On the nature of streaks in signal-detection. *Cognitive Psychology* 28:17–64. [aJDS]
Goldsmith, M. & Koriat, A. (1999) The strategic regulation of memory reporting: Mechanisms and performance consequences. In: *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application,* ed. D. Gopher & A. Koriat. MIT Press. [MG]
Goldsmith, M., Koriat, A. & Weinberg-Eliezer, A. (2002) The strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General* 131:73–95. [MG]
Gray, J. A. (1995) The contents of consciousness: A neuropsychological conjecture. *Behavioral and Brain Sciences* 18:659–722. [arJDS]
Green, D. & Swets, J. (1966/1974) *Signal detection theory and psychophysics,* 2nd edition. Krieger. [TON]
Grice, S., Spratling, M. W., Karmiloff-Smith, A., Halit, H., Csibra, G., de Haan, M. & Johnson, M. H. (2001) Disordered visual processing and oscillatory brain activity in autism and Williams syndrome. *NeuroReport* 12:2697–2700. [RC]
Griffin, D. R. (2001) *Animal minds, beyond cognition to consciousness.* University of Chicago Press. [DRG]
Hampton, R. R. (2001) Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences USA* 98:5359–62. [JC, RRH, JM, SJS, arJDS, LKS]
Hare, B., Call, J., Agnetta, B. & Tomasello, M. (2000) Chimpanzees know what conspecifics do and do not see. *Animal Behavior* 59:771–85. [VMW]
Hare, B., Call, J. & Tomasello, M. (2001) Do chimpanzees know what conspecifics know? *Animal Behavior* 61:139–51. [VMW]
Hart, J. T. (1965) Memory and the feeling-of-knowing experiments. *Journal of Educational Psychology* 57:347–49. [aJDS]
Hays, W. L. (1981) *Statistics.* CBS College. [aJDS]
Herbranson, W. T., Fremouw, T. & Shimp, C. P. (1999) The randomization procedure in the study of categorization of multi-dimensional stimuli by pigeons. *Journal of Experimental Psychology: Animal Behavior Processes* 25:113–35. [CPS, rJDS]
(2002) Categorizing a moving target in terms of its speed and direction. *Journal of the Experimental Analysis of Behavior* 78:249–70. [CPS]
Herman, L. M. (1975) Interference and auditory short-term memory in the bottlenosed dolphin. *Animal Learning and Behavior* 3:43–48. [aJDS]
Herman, L. M. & Arbeit, W. R. (1972) Frequency difference limens in the bottlenose dolphin: 1–70 kc/s. *Journal of Auditory Research* 2:109–20. [aJDS]
Herrnstein, R. J. (1990) Levels of stimulus control: A functional approach. *Cognition* 37:133–66. [aJDS]
Heyes, C. M. (1998) Theory of mind in nonhuman primates. *Behavioral and Brain Sciences* 21:101–48. [SJS, aJDS]
Heyes, C. M. & Huber, L., eds. (2000) *The evolution of cognition.* MIT Press. [DRG]
Higham, P. A. (2002) Strong cues are not necessarily weak: Thomson and Tulving (1970) and the encoding specificity principle revisited. *Memory and Cognition* 30:67–80. [MG]
Holmes, P. W. (1979) Transfer of matching performance in pigeons. *Journal of the Experimental Analysis of Behavior* 31:103–14. [aJDS]
Humphrey, N. K. (1976) The social function of the intellect. In: *Growing points in ethology,* ed. P. P. G. Bateson & R. A. Hinde. Cambridge University Press. [aJDS]
Inman, A. & Shettleworth, S. J. (1999) Detecting metamemory in nonverbal subjects: A test with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes* 25:389–95. [RRH, SJS, arJDS, TRZ]
James, W. (1890/1952) The principles of psychology. In: *Great books of the western world, vol. 53.* University of Chicago Press. [aJDS]
Jastrow, J. (1888) A critique of psycho-physic methods. *American Journal of Psychology* 1:271–309. [aJDS]
Karmiloff-Smith, A. (1992) *Beyond modularity: A developmental perspective on cognitive science.* MIT Press/Bradford Books. [JP, RC, rJDS]

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

(1998) Development itself is the key to understanding developmental disorders. *Trends in Cognitive Sciences* 2(10):389–98. [RC]
Karoly, P. (1993) Mechanisms of self-regulation: A systems view. *Annual Review of Psychology* 44:23–52. [aJDS]
Keenan, J. P., Ganis, G., Freund, S. & Pascual-Leone, A. (2000a) Self-face identification is increased with left hand responses. *Laterality* 5:259–68. [VMW]
Keenan, J. P., McCutcheon, B., Sanders, G., Freund, S., Gallup, G. G. & Pascual-Leone, A. (1999) Left hand advantage in a self- face recognition task. *Neuropsychologia* 37:1421–25. [VMW]
Keenan, J. P., Nelson, A., O'Connor, M. & Pascual-Leone, A. (2001) Self-recognition and the right hemisphere. *Nature* 409:305. [VMW]
Keenan, J. P., Wheeler, M. A., Gallup, G. G., Jr. & Pascual-Leone, A. (2000b) Self-recognition and the right prefrontal cortex. *Trends in Cognitive Science* 4:338–44. [VMW]
Köhler, W. (1925) *The mentality of apes.* Liveright. [DMR]
Koren, D., Seidman, L. J., Goldsmith, M., Poyurovsky, M., Zichel, S. & Klein, E. (2001) Insight and awareness of illness in first-episode schizophrenia: A metacognitive neuropsychological study. *Abstracts of the 8th Biennial International Congress on Schizophrenia Research, Schizophrenia Research* 49(supplement):112. [MG]
Koriat, A. (1993) How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review* 100:609–39. [JP, aJDS]
(1996) Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In: *Metacognition,* ed. J. Metcalfe & A. P. Shimamura. MIT Press. [RRH]
Koriat, A. & Goldsmith, M. (1994) Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General* 123:297–316. [MG]
(1996a) Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences* 19:167–88. [MG, rJDS, GM]
(1996b) Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review* 103:490–517. [MG]
Koriat, A., Goldsmith, M., Schneider, W. & Nakash-Dura, M. (2001) The credibility of children's testimony: Can children control the accuracy of their memory reports? *Journal of Experimental Child Psychology* 79:405–37. [MG]
Koriat, A. & Levy-Sadot, R. (2001) The combined contributions of the cue-familiarity and accessibility heuristics to feeling of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:34–53. [GM]
Kummer, H., Anzenberger, G. & Hemelrijk, C. K. (1996) Hiding and perspective taking in long-tailed macaques (*Macaca fascicularis*). *Journal of Comparative Psychology* 110:97–102. [SJS]
Laing, E., Hulme, C., Grant, J. & Karmiloff-Smith, A. (2001) Learning to read in Williams syndrome: Looking beneath the surface of atypical reading development. *Journal of Child Psychology and Psychiatry* 42(6):729–39. [RC]
Laursen, A. M. & Rasmussen, J. B. (1975) Circle-ellipse discrimination in man and monkey. *Vision Research* 15:173–74. [GM]
Lockhart, R. S. & Murdock, B. B. (1970) Memory and the theory of signal detection. *Psychological Bulletin* 74:100–109. [MG,TON]
Logue, A. W. (1988) Research on self-control: An integrating framework. *Behavioral and Brain Sciences* 11:665–709. [EF]
Lovibond, P. F. & Shanks, D. R. (2002) The role of awareness in Pavlovian conditioning: Empirical evidence and theoretical implications. *Journal of Experimental Psychology: Animal Behavior Processes* 28:3–26. [GM]
MacDonald, S. E. (1993) Delayed matching-to-successive-samples in pigeons: Short-term memory for item and order information. *Animal Learning and Behavior* 21:59–67. [SJS]
Mackintosh, N. J. (1971) Spatial reversal learning in rats, pigeons, and goldfish. *Psychonomic Science* 22:281–82. [TRZ]
MacMillan, N. A. & Creelman, C. D. (1991) *Detection theory: A user's guide.* Cambridge University Press. [SJS, arJDS]
Menzel, C. R. (1999) Unprompted recall and reporting of hidden objects by a chimpanzee (*Pan troglodytes*) after extended delays. *Journal of Comparative Psychology* 113:426–34. [arJDS]
Menzel, E. W., Savage-Rumbaugh, E. S. & Lawson, J. (1985) Chimpanzee (*Pan troglodytes*) spatial problem solving with the use of mirrors and televised equivalent mirrors. *Journal of Comparative Psychology* 99:211–17. [VMW]
Metcalfe, J. & Shimamura, A. (1994) *Metacognition: Knowing about knowing.* Bradford Books. [aJDS]
Miller, J. T., Saunders, S. S. & Bourland, G. (1980) The role of stimulus disparity in concurrently available reinforcement schedules. *Animal Learning and Behavior* 8:635–41. [aJDS]
Miller, R. R. & Oberling, P. (1998) Analogies between occasion setting and Pavlovian conditioning. In: *Occasion setting: Associative learning and cognition in animals,* ed. N. A. Schmajuk & P. C. Holland. APA Press. [GM]
Morgan, C. L. (1906) *An introduction to comparative psychology.* Walter Scott. [aJDS]
Morris, R. G. M. (2002) Episodic-like memory in animals: Psychological criteria, neural mechanisms and the value of episodic-like tasks to investigate models of neurodegenerative disease. In: *Episodic memory: New directions in research,* ed. A. Baddeley & J. P. Aggleton. Oxford University Press. [GM]
Nelson, T. O. (1987) The Goodman-Kruskal gamma coefficient as an alternative to signal-detection theory's measures of absolute-judgment accuracy. In: *Progress in mathematical psychology,* ed. E. Roskam & R. Suck. Elsevier Science. [TON]
ed. (1992) *Metacognition: Core readings.* Allyn and Bacon. [aJDS]
(1996) Consciousness and metacognition. *American Psychologist* 51:102–16. [aJDS]
Nelson, T. O. & Dunlosky, J. (1991) The delayed-JOL effect: When delaying your judgments of learning can improve the accuracy of your metacognitive monitoring. *Psychological Science* 2:267–70. [aJDS]
Nelson, T. O. & Narens, L. (1990) Metamemory: A theoretical framework and new findings. In: *The psychology of learning and motivation,* ed. G. Bower. Academic Press. [aJDS, LM, MG]
(1996) Why investigate metacognition? In: *Metacognition,* ed. J. Metcalfe & A. P. Shimamura. MIT Press. [RRH]
Pansky, A., Koriat, A., Goldsmith, M. & Pearlman-Avnion, S. (2002) Memory accuracy and distortion in old age: Cognitive, metacognitive, and neurocognitive determinants. Poster presented at the 30th Anniversary Conference of the National Institute for Psychobiology. [MG]
Parker, S. T., Mitchell, R. W. & Boccia, M. L., eds. (1994) *Self-awareness in animals and humans.* Cambridge University Press. [LM, aJDS]
Paterson, S. J., Brown, J. H., Gsödl, M. K., Johnson, M. H. & Karmiloff-Smith, A. (1999) Cognitive modularity and genetic disorders. *Science* 286(5448):2355–58. [RC]
Pavlov, I. P. (1932) Neuroses in man and animals. *Journal of the American Medical Association* 99:1012–13. [TRZ]
Perner, J. (1991) *Understanding the representational mind.* MIT Press. [RC]
(1998) The metaintentional nature of executive functions and theory of mind. In: *Language and thought,* ed. P. Carruthers & J. Boucher. Cambridge University Press. [JP]
Perner, J. & Wimmer, H. (1985) "John thinks that Mary thinks . . ." Attribution of second-order beliefs by 5- to 10-year-old children. *Journal of Experimental Child Psychology* 39 437–71. [VMW]
Piaget, J. (1952) *The origins of intelligence in children.* International University Press. [RC]
Platek, S. M., Critton, S. R., Myers, T. E. & Gallup, G. G., Jr. (in press) Contagious yawning: The role of self-awareness and mental state attribution. *Cognitive Brain Research.* [VMW]
Platek, S. M. & Gallup, G. G., Jr. (in press) Self-face recognition is affected by schizotypal personality traits. *Schizophrenia Research* 57:311–15. [VMW]
Platt, M. L., Brannon, E. M., Briese, T. L. & French, J. A. (1996) Differences in feeding ecology predict differences between golden lion tamarins (*Leontopithecus rosalia*) and Wied's marmoset (*Callithrix kuhli*) on spatial and visual memory tasks. *Animal Learning and Behavior* 24:384–93. [LKS]
Povinelli, D. (2000) *Folk physics for apes.* Oxford University Press. [JP]
Povinelli, D. J., Nelson, K. E. & Boysen, S. T. (1990) Inferences about guessing and knowing by chimpanzees (*Pan troglodytes*). *Journal of Comparative Psychology* 104:203–10. [JEK]
Povinelli, D. J., Parks, K. A. & Novak, M. A. (1991) Do rhesus monkeys (*Macaca mulatta*) attribute knowledge and ignorance to others? *Journal of Comparative Psychology* 105:318–25. [JEK]
Povinelli, D. J., Perilloux, H., Reaux, J. & Bierschwale, D. (1998) Young and juvenile chimpanzees' (*Pan troglodytes*) reactions to intentional versus accidental and inadvertent actions. *Behavioral Processes* 42:205–218. [VMW]
Povinelli, D. J., Rulf, A., Landau, K. & Bierschwale, D. (1993) Self-recognition in chimpanzees (Pan troglodytes): Distribution, ontogeny, and patterns of emergence. *Journal of Comparative Psychology* 107:347–72. [VMW]
Premack, D. (1978) On the abstractness of human concepts: Why it would be difficult to talk to a pigeon? In: *Cognitive processes in animal behavior,* ed. S. H. Hulse, H. Fowler & W. K. Honig. Erlbaum. [aJDS]
Premack, D. & Woodruff, G. (1978) Does the chimpanzee have a Theory of Mind? *Behavioral and Brain Sciences* (1):515–26. [PC]
Proust, J. (2001) A plea for mental acts. *Synthese* 129:105–28. [JP]
(submitted) Thinking of oneself. *Consciousness and Cognition.* [JP]
Purdy, J. E., Bales, S. L., Burns, M. L. & Wiegand, N. (1994) Assessing the rewarding aspects of a stimulus associated with extinction through the

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

observing response paradigm. *International Journal of Comparative Psychology* 7:101–16. [EF]
Purdy, J. E. & Peel, J. L. (1988) Observing response in goldfish (*Carassius auratus*). *Journal of Comparative Psychology* 102:160–68. [EF]
Rachlin, H. (1995) Self-control: Beyond commitment. *Behavioral and Brain Sciences* 18:109–59. [EF]
Ratcliff, R., Sheu, C. & Gronlund, S. (1992) Testing global memory models using ROC curves. *Psychological Review* 99:518–35. [TON]
Reder, L. M. (l987) Selection strategies in question answering. *Cognitive Psychology* 19:90–138. [LKS]
ed. (1996) *Implicit memory and metacognition.* Erlbaum. [JHF, GM, aJDS]
Reder, L. M. & Shunn, C. D. (1996) Metacognition does not imply awareness: Strategy choice is governed by implicit learning and memory. In: *Implicit memory and metacognition,* ed. L. M. Reder. Erlbaum. [GM, JP, LKS]
Reiss, D. & Marino, L. (2001a) Mirror self-recognition in the bottlenose dolphin: A case of cognitive convergence. (August 12, 2001). http://www.emory.edu/LIVING_LINKS/dolphin_MSR.pdf [VMW]
(2001b) Self-recognition in the bottlenose dolphin: A case of cognitive convergence. *Proceedings of the National Academy of Sciences USA* 98:5937–42. [LM,rJDS]
Roberts, W. A. & Kraemer, P. J. (1981) Recognition memory for lists of visual stimuli in monkeys and humans. *Animal Learning and Behavior* 9:587–94. [aJDS]
Roitblat, H. L., Penner, R. H. & Nachtigall, P. E. (1990) Matching-to-sample by an echolocating dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: Animal Behavior Processes* 16:85–95. [rJDS]
Rosellini, R. A. & Terris W. (1976) Fear as a discriminative stimulus for an appetitive instrumental response. *Learning and Motivation* 7:3127–339. [JEK]
Rozin, P. (1976) The evolution of intelligence and access to the cognitive unconscious. *Progress in Psychobiology and Physiological Psychology* 6:245–80. [rJDS]
Rumbaugh, D. M. (2002) Emergents and rational behaviorism. *Eye on Psi Chi* 6:8–14. [DMR]
Rumbaugh, D. M., Savage-Rumbaugh, E. S. & Washburn, D. A. (1996a) Toward a new outlook on primate learning and behavior: Complex learning and emergent processes in comparative perspective. *Japanese Psychological Research* 38:113–25. [DMR]
Rumbaugh, D. M., Washburn, D. A. & Hillix, W. A. (1996b) Respondents, operants, and emergents: Toward an integrated perspective on behavior. In: *Learning as a self-organizing process,* ed. K. Pribram & J. King. Erlbaum. [DMR]
Sands, S. F. & Wright, A. A. (1980) Primate memory: Retention of serial list items by a rhesus monkey. *Science* 209:938–39. [aJDS]
Schrier, A. M. & Wing, T. G. (1973) Eye movements of monkeys during brightness discrimination and discrimination reversal. *Animal Learning and Behavior* 1:145–50. [rJDS]
Schull, J. & Smith, J. D. (1992) Knowing thyself, knowing the other: They're not the same. *Behavioral and Brain Sciences* 15:166–67. [aJDS]
Schusterman, R. J. & Barrett, B. (1975) Detection of underwater signals by a California sea lion and a bottlenose porpoise: Variation in the payoff matrix. *Journal of the Acoustical Society of America* 57:1526–37. [aJDS]
Schwartz, B. L. (1994) Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin and Review* 1:357–75. [aJDS]
Schwartz, B. L., Colon, M. R., Sanchez, I. C., Rodriguez, I. A. & Evans, S. (2002) Single-trial learning of "what" and "who" information in a gorilla (*Gorilla gorilla gorilla*): Implications for episodic memory. *Animal Cognition* 5:85–90. [arJDS]
Shallice, T. & Burgess, P. (1991) Higher-order cognitive impairments and frontal lobe lesions in man. In: *Frontal lobe function and dysfunction,* ed. H. S. Levin, H. M. Eisenberg & A. L. Benton. Oxford University Press. [JP]
Shettleworth, S. J. (1998) *Cognition, evolution, and behavior.* Oxford University Press. [SJS]
Shields, W. E. (1999) *Nonverbal judgments of knowing by humans and rhesus monkeys.* Doctoral dissertation, State University of New York at Buffalo. [MG, JM, aJDS]
Shields, W. E., Smith, J. D., Guttmannova, K. & Washburn, D. A. (2002) A study of retrospective confidence judgments by rhesus monkeys. (manuscript in preparation.) [aJDS]
Shields, W. E., Smith, J. D. & Washburn, D. A. (1997) Uncertain responses by humans and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task. *Journal of Experimental Psychology: General* 126:147–64. [arJDS]
Shiffrin, R. M. & Schneider, W. (1977) Controlled and automatic human information processing: Perceptual learning, automatic attending and a general theory. *Psychological Review* 84:127–90. [JP,aJDS]
Shillito, D. J. (2002) An examination of the cognitive precursors to a mature theory of mind in orangutans. Unpublished doctoral dissertation, SUNY-Albany, Dept. of Psychology. [VMW]
Shillito, D. J., Gallup, G. G., Jr. & Beck, B. B. (1999) Factors affecting mirror behavior in western lowland gorillas, *Gorilla gorilla. Animal Behavior* 57:999–1004. [VMW]
Shimp, C. P. (1982) Metaknowledge in the pigeon: An organism's knowledge about its own behavior. *Animal Learning and Behavior* 10:358–64. [CPS, rJDS]
(1983) The local organization of behavior: Dissociations between a pigeon's behavior and self-reports of that behavior. *Journal of the Experimental Analysis of Behavior* 39:61–68. [CPS, rJDS]
(1999) Tolerance in a rigorous science. *Journal of the Experimental Analysis of Behavior* 71:284–88. [CPS]
(2001) Behavior as a social construction. *Behavioral Processes* 54:11–32. [CPS]
(in press) A case study in scientific peer review: Local and global analyses. *Journal of the Experimental Analysis of Behavior.* [CPS]
Smith, J. D. (1995) The homunculus at home. Commentary on J. A. Gray, The contents of consciousness: A neuropsychological conjecture. *Behavioral and Brain Sciences* 18:697–98. [arJDS]
(2002) Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science* 13:437–42. [rJDS]
Smith, J. D. & Minda, J. P. (1998) Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 24:1411–36. [arJDS]
(2000) Thirty categorization results in search of a model. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 26:3–27. [aJDS]
(2001) Journey to the center of the category: The dissociation in amnesia between categorization and recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 27:984–1002. [rJDS]
Smith, J. D. & Schull, J. (1989) A failure of uncertainty monitoring in the rat. (unpublished data). [aJDS]
Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R. & Erb, L. (1995) The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General* 124:391–408. [MG, aJDS]
Smith, J. D., Shields, W. E., Allendoerfer, K. R. & Washburn, D. A. (1998) Memory monitoring by animals and humans. *Journal of Experimental Psychology: General* 127:227–50. [MG, arJDS]
Smith, J. D., Shields, W. E., Schull, J. & Washburn, D. A. (1997) The uncertain response in humans and animals. *Cognition* 62:75–97. [SJS, arJDS, TRZ]
Smith, S. M., Brown, J. M. & Balfour, S. P. (1991) TOTimals: A controlled experimental method for studying tip-of-the-tongue states. *Bulletin of the Psychonomic Society* 29:445–47. [aJDS]
Sole, L. M., Shettleworth, S. J. & Bennett, P. J. (in press) Uncertainty in pigeons. *Psychonomic Bulletin and Review.* [SJS, rJDS]
Son, L. K. & Kornell, N. (in press) Confidence judgments in rhesus macaques: Does self-awareness follow? In: *The missing link in cognition: Origins of self-knowing consciousness,* ed. H. S. Terrace & J. Metcalfe. Oxford University Press. [rJDS, LKS]
Son, L. K., Kornell, N. & Terrace, H. S. (2002) Confidence judgments in rhesus macaques. Poster presented at the Annual Meeting of the Psychonomic Society. [LKS]
Spehn, M. K. & Reder, L. M. (2000) The unconscious feeling of knowing: A commentary on Koriat's paper. *Consciousness and Cognition* 9:187–92. [JP]
Stuss, D. T., Gallup, G. G., Jr. & Alexander, M. P. (2001) The frontal lobes are necessary for theory of mind. *Brain* 124:279–86. [VMW]
Suddendorf, T. & Whiten, A. (2001) Mental evolution and development: Evidence for secondary representation in children, great apes, and other animals. *Psychological Bulletin* 127:629–50. [JC]
Swartz, K. B. (1997) What is mirror self-recognition in nonhuman primates, and what is it not? In: *The self across psychology: Self-recognition, self-awareness, and the self concept. Annals of the New York Academy of Sciences* 818:65–71, ed. J. G. Snodgrass & R. L. Thompson. New York Academy of Sciences. [aJDS]
Swartz, K. B., Sarauw, D. & Evans, S. (1999) Comparative aspects of mirror self-recognition in great apes. In: *The mentalities of gorillas and orangutans: Comparative perspectives,* ed. S. T. Parker, R. W. Mitchell & M. L. Boccia. Cambridge University Press. [aJDS]
Swets, J. A., Tanner, W. P. & Birdsall, T. G. (1961) Decision processes in perception. *Psychological Review* 68:301–40. [aJDS]
Suarez, S. D. & Gallup, G. G., Jr. (1981) Self-recognition in chimpanzees and orangutans, but not gorillas. *Journal of Human Evolution* 10:175–88. [VMW]
Teller, S. A. (1989) Metamemory in the pigeon: Prediction of performance on a delayed matching to sample task. Unpublished undergraduate thesis, Reed College. [arJDS]
Terman, M. & Terman, J. (1972) Concurrent variation of response bias and

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press

sensitivity in an operant-psychophysical test. *Perception and Psychophysics* 11:428–32. [aJDS]

Terrace, H. & Metcalfe, J. (in press) *The evolution of self-reflective consciousness.* Oxford University Press. [JM]

Terrace, H. S., Son, L. & Brannon, E. M. (2003) Serial expertise in rhesus macaques. *Psychological Science* 14:66–73. [GM]

Thompson, R. A. (1998) Early sociopersonality development. In: *Handbook of child psychology, vol. 3,* ed. N. Eisenberg. Series: *Social, emotional, and personality development, 5th edition,* ed. W. Damon (Series editor). Wiley. [JHF]

Thomson, G. H. (1920) A new point of view in the interpretation of threshold measurements in psychophysics. *Psychological Review* 27:300–307. [aJDS]

Tolman, E. C. (1927) A behaviorist's definition of consciousness. *Psychological Review* 34:433–39. [aJDS]

(1932/1967) *Purposive behavior in animals and men.* The Century Company. [aJDS]

(1938) The determiners of behavior at a choice point. *Psychological Review* 45:1–41. [aJDS, TRZ]

Tomasello, M. & Call, J. (1997) *Primate cognition.* Oxford University Press. [JC, JEK]

Tomasello, M., Call, J. & Hare, B. (1998) Five primate species follow the visual gaze of conspecifics. *Animal Behavior* 55:1063–69. [VMW]

Tomasello, M., Hare, B. & Fogleman, T. (2001) The ontogeny of gaze following in chimpanzees, *Pan troglodytes,* and rhesus macaques, *Macaca mulatto. Animal Behaviour* 61:335–43. [VMW]

Treisman, M. & Faulkner, A. (1984) The setting and maintenance of criteria representing levels of confidence. *Journal of Experimental Psychology: Human Perception and Performance* 10:119–39. [aJDS]

Urban, F. M. (1910) The method of constant stimuli and its generalizations. *Psychological Review* 17:229–59. [aJDS]

Vogeley, K., Bussfeld, P., Newen, A., Herrmann, S., Happe, F., Falkai, P., Maier, W., Shah, N. J., Fink, G. R. & Zilles, K. (2001) Mind reading: Neural mechanisms of theory of mind and self-perspective. *NeuroImage* 14(1 Pt 1):170–81. [VMW]

Washburn, D. A. & Rumbaugh, D. M. (1992) Testing primates with joystick-based automated apparatus: Lessons from the Language Research Center's Computerized Test System. *Behavior Research Methods, Instruments, and Computers* 24:157–64. [aJDS]

Washburn, D. A., Smith, J. D., Baker, L. A. & Raby, P. R. (2001) Responding to uncertainty: Individual differences and training effects. *Proceedings of the 2001 meeting of the Human Factors and Ergonomics Society* 911–15. [aJDS]

Watson, C. S., Kellogg, S. C., Kawanishi, D. T. & Lucas, P. A. (1973) The uncertain response in detection-oriented psychophysics. *Journal of Experimental Psychology* 99:180–85. [aJDS]

Weaver, J. E., Steirn, J. N. & Zentall, T. R. (1997) Transitive inference in pigeons: Control for differential value transfer. *Psychonomic Bulletin and Review* 4:113–17. [TRZ]

Weiskrantz, L. (1986) *Blindsight: A case study and implications.* Oxford University Press. [aJDS]

(1997) *Consciousness lost and found: A neuropsychological exploration.* Oxford University Press. [aJDS]

Whiten, A. & Byrne, R. W., eds. (1997) *Machiavellian intelligence II: Extensions and evaluations.* Cambridge University Press. [aJDS]

Wilson, T. (2002) *Strangers to ourselves.* Harvard University Press. [PC]

Woodworth, R. S. (1938) *Experimental psychology.* Holt. [aJDS]

Wright, A. A., Santiago, H. C., Sands, S. F., Kendrick, D. F. & Cook, R. G. (1985) Memory processing of serial lists by pigeons, monkeys, and people. *Science* 229:287–89. [aJDS]

Wright, A. A., Shyan, M. R. & Jitsumori, M. (1990) Auditory same/different concept learning by monkeys. *Animal Learning and Behavior* 18:287–94. [aJDS]

Wyckoff, L. B., Jr. (1952) The role of observing responses in discrimination learning: Part 1. *Psychological Review* 59:431–42. [EF]

Yunker, M. P. & Herman, L. M. (1974) Discrimination of auditory temporal differences by the bottlenose dolphin and by the human. *Journal of the Acoustical Society of America* 56:1870–75. [aJDS]

Zentall, T. R. (1998) Symbolic representation in pigeons: Emergent stimulus relations in conditional discrimination learning. *Animal Learning and Behavior* 26:363–77. [TRZ]

Zentall, T. R., Clement, R. S., Bhatt, R. S. & Allen, J. (2001) Episodic-like memory in pigeons. *Psychonomic Bulletin and Review* 8:685–90. [ZPS,TRZ]

Zentall, T. R., Sutton, J. E. & Sherburne, L. M. (1996) True imitative learning in pigeons. *Psychological Science* 7:343–46. [TRZ]

https://doi.org/10.1017/S0140525X03220082 Published online by Cambridge University Press