

Animal Metacognition: A Tale of Two Comparative Psychologies

J. David Smith

University at Buffalo, The State University of New York

Justin J. Couchman

Fredonia State University

Michael J. Beran

Georgia State University

A growing literature considers whether animals have capacities that are akin to human metacognition (i.e., humans' capacity to monitor their states of uncertainty and knowing). Comparative psychologists have approached this question by testing a dolphin, pigeons, rats, monkeys, and apes using perception, memory, and food-concealment paradigms. As part of this consideration, some associative modelers have attempted to describe animals' "metacognitive" performances in low-level, associative terms—an important goal if achievable. The authors summarize the empirical and theoretical situation regarding these associative descriptions. The associative descriptions in the animal-metacognition literature fail to encompass important phenomena. The sharp focus on abstract, mathematical associative models creates serious interpretative problems. The authors compare these failed associative descriptions with an alternative theoretical approach within contemporary comparative psychology. The alternative approach has the potential to strengthen comparative psychology as an empirical science and integrate it more fully within the mainstream of experimental psychology and cognitive science.

Keywords: metacognition, comparative psychology, metamemory, primate cognition, associative learning

Metacognition is the capacity to monitor and control one's cognitive processes. The idea in the metacognition research literature is that humans have a cognitive executive that oversees and guides—and thereby optimizes—cognition (Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Bjork, 2008; Flavell, 1979; Koriat, 1993; Metcalfe & Shimamura, 1994; Nelson, 1992; Schwartz, 1994; Serra & Dunlosky, 2005). These regulatory functions are assessed in the laboratory by collecting metacognitive judgments—feelings of knowing, tip-of-the-tongue states, confidence judgments, and so forth. Metacognition is crucial to humans' learning, thinking, comprehension, and education. It is one of humans' sophisticated cognitive capacities (Metcalfe & Kober, 2005). It emerges late in their cognitive development (Balcomb & Gerken, 2008). It is linked to their declarative cognitive system (Koriat, 2007; Nelson, 1996) and self-awareness (Gallup, 1982). For all these reasons, one might suppose that metacognition is a uniquely human cognitive capacity.

Naturally, therefore, researchers have begun to ask whether other species have a cognitive capacity that is a functional analog

to human metacognition. The answer could reveal the phylogenetic roots of human metacognition, suggest the earliest forms of metacognition in human cognitive development, and provide animal models for studying and fostering human metacognition. It could also illuminate animals' reflective minds and their cognitive self-awareness. Animal metacognition is now an influential area within comparative psychology (reviews in Hampton, 2009; Smith, Beran, & Couchman, 2012; Kornell, 2009; Smith, Couchman, & Beran, 2012; Smith, Countinho, Boomer, & Beran, 2012).

Beran, Smith, and Perdue (2013) typifies animals' success in uncertainty-monitoring tasks. They tested language-trained chimpanzees in an information-seeking task. A food item was within an opaque container. Sometimes the chimpanzees saw it when it was placed, sometimes not. They were given the food if they named the item on their lexigram symbol keyboard. Chimpanzees were significantly more likely to visit the food container first—before choosing a lexigram—on trials in which they did not know the container's contents. They were likely to just name the item on the keyboard—without looking into the container—on trials in which they had earlier seen the contents of the container. Thus, chimpanzees showed efficient information-seeking behavior that suggested they knew whether they knew the item's identity when it was time to name it (see also Call, 2010; Call & Carpenter, 2001; Suda-King, 2008).

Summarizing across many studies, it appears that animals' uncertainty responses—given their flexibility, generalizability, and so forth—deserve a higher-level cognitive interpretation that rises above interpretations based in stimulus cues and reinforcement history. For example, animals show adaptive uncertainty responses facing abstract conceptual judgments (Shields, Smith, & Washburn, 1997) and indeterminate memories (Hampton, 2001; Smith,

This article was published Online First August 19, 2013.

J. David Smith, Department of Psychology and Center for Cognitive Science, University at Buffalo, The State University of New York; Justin J. Couchman, Department of Psychology, Fredonia State University; Michael J. Beran, Language Research Center, Georgia State University.

The preparation of this article was supported by Grant 1R01HD061455 from NICHD and Grant BCS-0956993 from NSF.

Correspondence concerning this article should be addressed to J. David Smith, 346 Park Hall, Department of Psychology, SUNY Buffalo, Buffalo, NY 14260. E-mail: psysmith@buffalo.edu

Shields, Allendoerfer, & Washburn, 1998; Kornell, Son, & Terrace, 2007). They can respond uncertain absent any cuing stimuli (Hampton, 2001), on the first trial of novel tasks (Washburn, Smith, & Shields, 2006), and even when there are no immediate feedback or reinforcement signals so that the processes of conditioning and association are ruled out (Smith, Beran, Redford, & Washburn, 2006).

Based on these and many other results (e.g., Beran & Smith, 2011; Kornell et al., 2007; Washburn, Gullledge, Beran, & Smith, 2010), there is some consensus that some animals can show uncertainty processes that are functionally equivalent to human metacognition. Sutton and Shettleworth (2008, p. 266) concluded that “metamemory, the ability to report on memory strength, is clearly established in rhesus macaques (*Macaca mulatta*) by converging evidence from several paradigms.” Fujita (2009, Abstract) concluded that “evidence for metacognition by nonhuman primates has been obtained in great apes and old world monkeys.” Roberts, Feeney, McMillan, MacPherson, and Musolino (2009, p. 130) concluded that “substantial evidence from several laboratories converges on the conclusion that rhesus monkeys show metacognition in experiments that require behavioral responses to cues that act as feeling of knowing and memory confidence judgments.”

However, there are some who still emphasize the need to remain true to Morgan’s (1906) canon and describe animals’ behavior at the lowest psychological level. Even given seemingly metacognitive performances by some animals, they point out that one must consider explaining those performances using associative mechanisms. That is, those performances might have low-level and nonmetacognitive bases if animals’ uncertainty responses were really cued by stimuli or conditioned/trained by reinforcement contingencies. This is a constructive concern that has attended and influenced comparative metacognition research for a decade (e.g., Carruthers, 2008; Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009a, 2009b; Le Pelley, 2012; Smith, Beran, Couchman, Coutinho, & Boomer, 2009a, 2009b; Smith, Beran, Coutinho, & Couchman, 2008; Staddon, Jozefowicz, & Cerutti, 2007).

The concern can be illustrated using a common uncertainty-monitoring task. Uncertainty tasks are often organized along stimulus continua. For example, stimulus rectangles—to be classified as Sparse or Dense—might vary along a 41-step continuum in their density of lit pixels. Levels 1–20 would be rewarded as Sparse; Levels 22–41 as Dense; Level 21—the exact breakpoint of the discrimination—would not be presented. Animals would have two primary discrimination responses with which to classify stimuli as Sparse or Dense. For correct responses, they would earn a food reward. For incorrect responses, they would earn a trial-less timeout period during which no rewards could be earned. They would also be given an “uncertainty response” (UR) with which to decline to complete any trials they choose. The UR would simply remove the present trial and move the animal on to the next randomly selected trial. It would never bring any direct reward or other consequence.

The trials near the discrimination breakpoint (Level 21) will be objectively difficult, causing errors and reducing rewards. Ideally, subjects would choose to complete easy trials (near the ends of the stimulus continuum) by making sparse and dense responses, and would reserve the UR for the difficult trial levels near the breakpoint. Humans have shown this pattern repeatedly, and they report being consciously uncertain when they make URs. Macaques have

also shown this pattern repeatedly, though they report nothing. These empirical results are not disputed.

The theoretical dispute is the UR’s psychological interpretation. It might reflect the animal’s judging difficulty or monitoring uncertainty—capacities analogous to metacognition. It might reflect low-level reactions based in aversion or avoidance. That is, animals might have stored the task’s reinforcement history, together with behavioral tendencies that dictate that poorly rewarded, often punished middle stimuli are aversive and should be avoided using the default avoidance response (UR). By this interpretation, the UR would really just be a “Middle” response (MR), entrained just as Sparse and Dense responses are, subject to reinforcement pressures and low-level associative-learning mechanisms just as Sparse and Dense responses are.

Pursuing these low-level interpretations, some researchers have constructed and fit a variety of associative models to the uncertainty-monitoring data produced by animals. For example, Smith, Beran, Coutinho, and Couchman (2008) developed an associative model based in signal-detection theory. It assumed that animals store the reinforcement histories associated with different stimuli and that they respond with aversion-avoidance responses (not uncertainty responses) when they encounter error-causing stimuli. They found that the associative model could fit the data from rhesus macaques in a Sparse–Dense discrimination like that just described and used in Smith, Shields, Schull, and Washburn (1997). Similarly, Staddon, Jozefowicz, and Cerutti (2007, also Jozefowicz et al., 2009a, 2009b) used a closely related signal-detection model—called the Behavioral Economic Model—to fit the uncertainty-monitoring data produced by rats in a temporal-discrimination task used in Foote and Crystal (2007). Most recently, Le Pelley (2012) asked whether an associative-learning model could simulate the way that animals learn trial-by-trial to perform adaptively in uncertainty-monitoring tasks. He also assumed that there were stimulus and response registers that were gradually updated over trials by reinforcement, so that adaptive behavioral choices gradually emerged through the processes of associative learning.

The present article summarizes the empirical and theoretical situation regarding these associative models and descriptions of animal-metacognition phenomena. First, we consider four of the field’s phenomena. In each area, the associative descriptions in the animal-metacognition literature fail to capture the crucial phenomena, and indeed those descriptions may be conceptually and theoretically misguided. Then, we consider at a broader theoretical level why the concerted attempt to produce an “associative” account in this empirical domain is flawed and carries risks to the ongoing theoretical development of comparative psychology. Next, we show how understanding these risks informs one’s interpretation of cross-species research in metacognition. Finally, we contrast two different approaches to interpretation within this area of comparative psychology, one offered by a group of determined associative theorists and one by other animal-metacognition researchers. In our view, the latter approach shows its own distinctive promise to strengthen interpretation in comparative psychology broadly, to broaden comparative psychology’s reach, to produce a constructive dialog between comparative and cognitive psychologists, and to integrate comparative psychology more fully into the mainstream of experimental psychology and cognitive science.

Dissociating Uncertainty From Reinforcement (Smith et al., 2006)

Synopsis of Smith, Beran, Redford, and Washburn (2006)

Associative descriptions of URs are based in stimuli, reinforcement, and punishment. They depend on transparent feedback to entrain reactive URs when stimuli signal likely punishment or unlikely reward. To test these descriptions, Smith et al. (2006, also Couchman, Coutinho, Beran, & Smith, 2010) replaced trial-by-trial feedback with deferred feedback whereby macaques worked for four trials before receiving any feedback. At that point, macaques received together all of their rewards for correct responses and then all of their error timeouts. Now monkeys could not associate specific responses or stimulus-response pairs with specific outcomes. In particular, they could not associate the UR with difficult trial levels, because they never received any direct or immediate feedback or outcomes for responses made to those stimuli. For this reason, the normal processes of associative learning were disrupted. They had no basis for avoiding nonreinforcing stimuli—that basis was denied them. But they could still potentially monitor difficulty, or feel uncertainty, and so forth. Accordingly, this paradigm created the interesting possibility that one might observe an animal respond to its uncertainty about trial levels it performed highly accurately, or respond confidently about trial levels it very often classified incorrectly. Both findings would undermine associative descriptions.

Figure 1A confirms that these undermining findings did emerge in the data from one of Smith et al.'s (2006) two macaques. At Level 9, macaque Murph (in a Sparse–Dense discrimination task like that described earlier) made many URs (23%) for trials he answered 97% correct (when he tried them on other occasions). By any associative model, if he had tracked the reinforcement history for these trials (despite the deferred–rearranged feedback), he would have eagerly accepted these objectively easy, winning, rewarding trials. At Level 20, he made fewer URs (19%) for trials he answered 22% correct (when he tried them). Note that he was far below chance accuracy on these trials. If he had tracked the reinforcement history for these trials (despite the deferred–rearranged feedback), he would have urgently declined these disastrous, losing trials and saved himself many seconds of penalty timeouts.

However, the two trial types felt the same to him in the sense of being about equally worth trying. This suggests that the paradigm successfully camouflaged the task's associative structure. Indeed, Figure 1B confirms that over most of the range of performance accuracy there was no relationship between the proportion of correct responses and the proportion of URs. No associative model would predict this independence—URs (taken as associative aversion responses) should condition to the aversive, losing trials. If the associative structure of the task was camouflaged, then associative-learning processes probably cannot have been at work exclusively or even predominantly in this monkey's performance. And perhaps they demonstrably were not at work because the animal used the UR in the same way for stimuli that varied in their reinforcement likelihood by more than a factor of four.

Interpretation

For these reasons, the findings in Smith et al. undermine associative descriptions, and so the stakes are high for associative models to try to explain them. Neither Staddon et al. (2007) nor Jozefowicz et al., (2009a, 2009b) tried to do so. Their model depends on the UR's receiving a small reward giving it positive associative strength. But URs in Smith et al. (2006) received no direct reinforcement so this associative model cannot capture the result. The Staddon/Jozefowicz model also depends on animals' trial–outcome experience so that they can associate different stimuli with different payoff probabilities. But the deferred–reinforcement denied Murph this experience. He had no way to form these associations.

Smith et al. (2008) did try to fit a related associative model to data in Smith et al. (2006). But the model's fit was very poor. The error of prediction between predicted and observed response proportions was 5%, and this is large in almost all formal-analytic domains (e.g., Smith, Coutinho, Church, & Beran, 2013; Smith, Redford, & Haas, 2008). From Levels 1–31, the theoretically important UR was mispredicted by 7% per stimulus level on average.

Smith et al.'s (2008) associative model failed for a fundamentally important reason. Associative models are anchored to the task's category definitions and reinforcement structure. Dense responses on Dense trials and Sparse responses on Sparse trials earn rewards. Therefore, associative response curves essentially must be centered on the task's true discrimination breakpoint, and this is how Smith et al.'s associative model behaved. But in sharp contrast, the macaque's response curves were displaced to the left from the task's reinforcement structure centered at Trial Level 21. Associative models, as Smith et al. found, and as we will see again momentarily, do not capture this displacement. The displacement actually confirms that the animal ignored the task's associative and reinforcement signals, or perhaps that the animal could not detect them because of the deferred–rearranged feedback.

Le Pelley (2012) tried to apply an associative-learning model to Smith et al.'s (2006) data. To illustrate the application of associative models to animal-metacognition research, the application of Le Pelley's model to Smith et al.'s data is discussed in more depth in the Appendix. The question is whether it can reproduce the displaced UR curves that Smith et al. found. We programmed Le Pelley's model to answer this question. Figure 2A shows the UR curves by simulated performers in Le Pelley's model that did show a leftward displacement of that curve. The UR curves only shift leftward as the area under the UR curve grows—that is, as the UR's response dominance grows. At the extreme, the UR runs amok to control fully half the stimulus environment, which would cost a real animal half its task rewards. Consequently, Le Pelley's model never produces anything like the UR data pattern that Smith et al. observed (Figure 1C). Figure 5 in Le Pelley makes it clear that there is no disagreement between our and his computational model. In the nine simulations displayed of Smith et al.'s paradigm, there is the same overall tendency exhibited (i.e., the asymmetry of UR responding away from Level 21 is always caused by broadly dominant uncertainty responding).

Confirming the point about broadening and dominance, Figure 2B shows—for 72 runs of Le Pelley's simulation—the mean level of URs (averaged across 40 stimulus levels) plotted against the

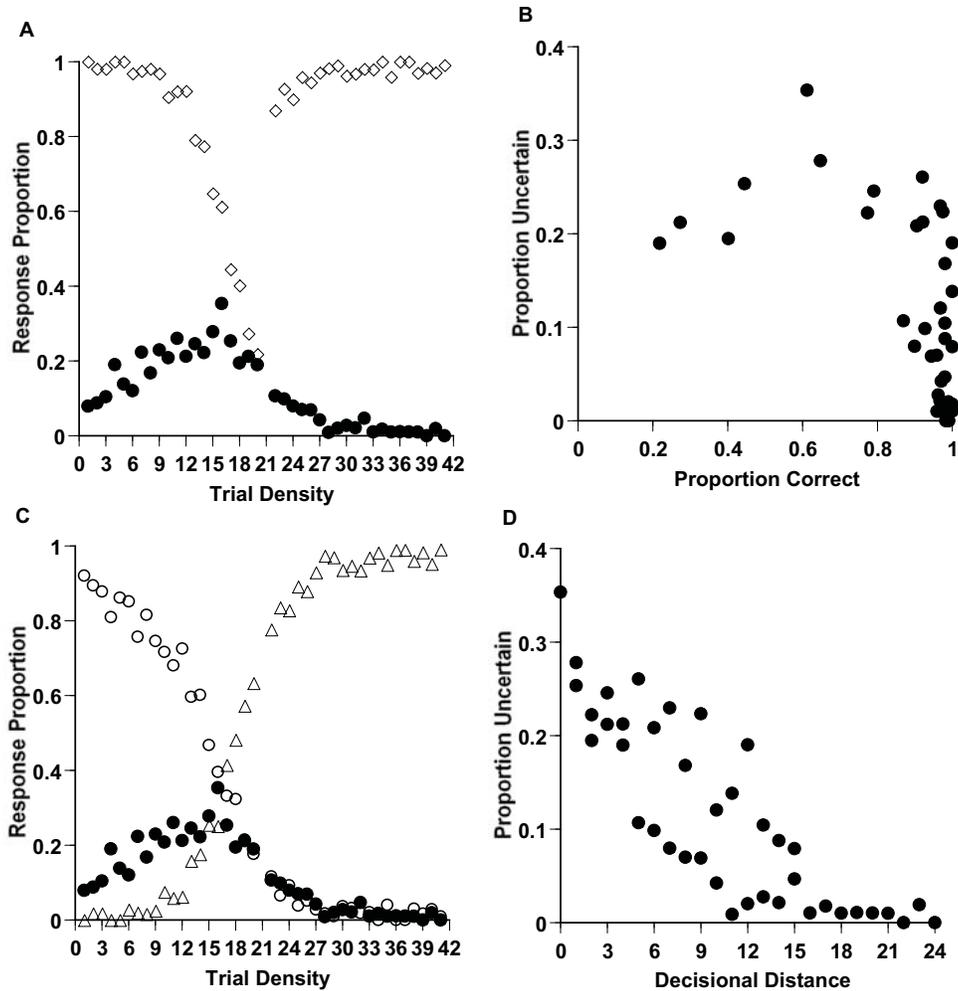


Figure 1. A macaque's performance in Smith et al.'s (2006) Sparse–Dense discrimination. A. The horizontal axis indicates the density level of the box. The Sparse and Dense responses, respectively, were correct for boxes at density Levels 1–20 and 22–41. The open diamonds show the proportion of trials attempted that were answered correctly. The dark circles show the proportion of trials receiving the uncertainty response at each density level. B. The macaque's performance in the same task, with the proportion of trials declined at each trial level plotted against the proportion of correct responses. C. The macaque's performance showing separately his use of the Sparse and Dense responses (open diamonds and open triangles). D. The macaque's performance showing his proportion of trials declined at each trial level plotted against the decisional distance of the level from his decisional breakpoint (Level 16 = 0; Levels 15 and 17 = 1; etc.). From *Dissociating Uncertainty States and Reinforcement Signals in the Comparative Study of Metacognition*, by J. D. Smith, M. J. Beran, J. S. Redford, and D. A. Washburn, 2006, *Journal of Experimental Psychology: General*, 135, p. 292. Copyright, 2006 by the American Psychological Association. Reprinted with permission.

fulcrum point of URs. This fulcrum point is the median trial level of URs—that is, the density level at which half the simulated observer's URs lay above that level and half below that level. This fulcrum reflects symmetry or asymmetry in URs. At low levels of URs, the simulation is symmetrical with URs clustering near the discrimination breakpoint at Level 21. Appropriately, the simulation aligns with the task's associative structure. In a rough V shape, URs only become asymmetrical when their overall level sharply increases. The real macaque's performance point in this space is 14 and 0.12. The simulation framework presents no values near that place. Like flies in a bottle, the associative solutions to

the task afforded in this model are constrained in the performance patterns they can produce and the points representing them are bounded in the positions in performance space they can occupy, and they cannot reach out to where the macaque's nonassociative performance lay.

We also fit these runs of the model to Smith et al.'s (2006) data. That is, we found how close each set of 120 simulated performance levels (three response types at 40 stimulus levels) came to matching the corresponding 120 observed performance levels in Smith et al. To summarize the fits, we used the Average Absolute Deviation (AAD) that represents the average

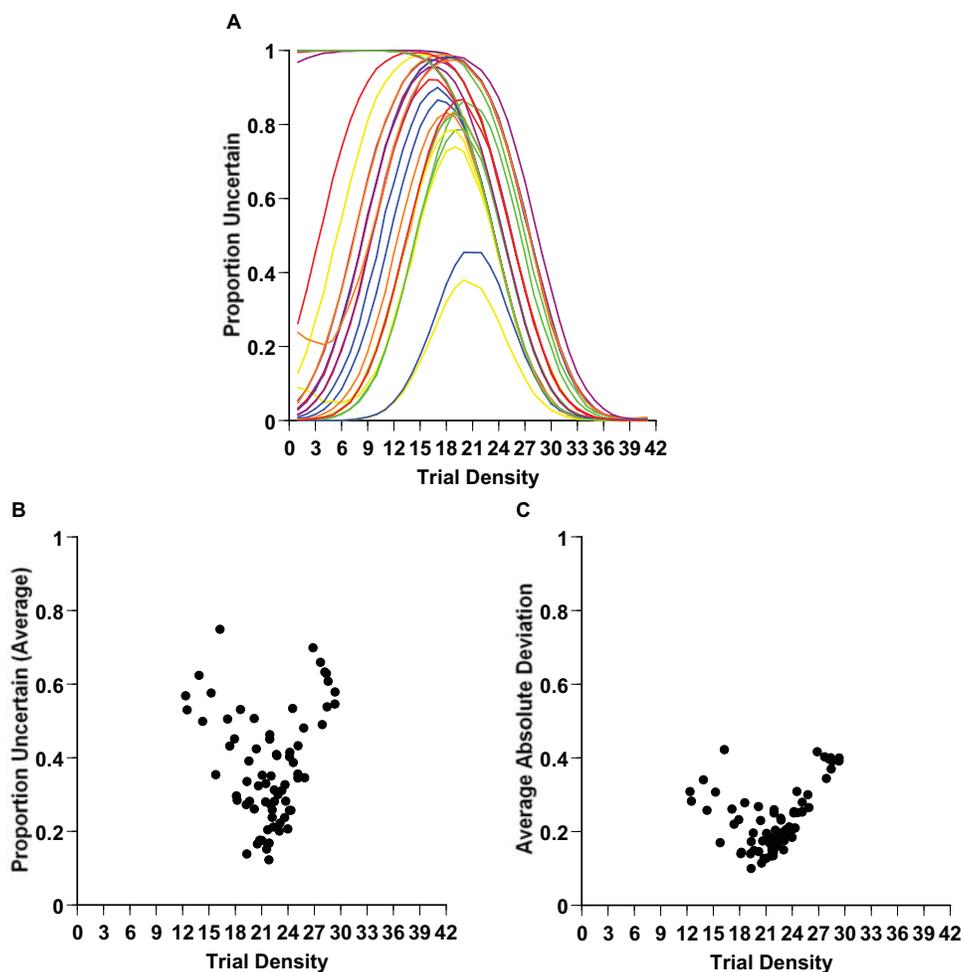


Figure 2. A. The UR (uncertainty response) curves of simulated performers in Le Pelley's (2012) associative model. B. The mean proportion of URs (averaged across trial levels) plotted against the fulcrum of URs (i.e., the point on the perceptual continuum that placed half of the URs to either side of it). C. The Average Absolute Deviation (a measure described in the text that summarizes the fit of simulated performers' data pattern to the data in Smith et al., 2006) plotted against the fulcrum of URs.

of the 120 errors of prediction, with all errors of prediction taken positively. Figure 2C shows this AAD fit measure plotted against the fulcrum of URs. The associative model at best (i.e., at the minimum AAD) averaged a misprediction of .10 for each of 120 data points. This poor fit by itself sharply undermines the descriptive value of Le Pelley's associative-learning model in this case. At a UR fulcrum of 14, shown by the real macaque, Le Pelley's (2012) associative model would miss the real data by about .30 per data point—an extremely poor fit. This fit would mean that an observed .2 proportion of sparse responses for a trial level would be mispredicted as .5; a .8 observed proportion of sparse responses for a trial level would also be mispredicted as .5. One could not even differentiate response proportions of .2 and .8. These are fits with which no formal modeler would rest comfortably.

In contrast, Smith et al. (2006) suggested a simple nonassociative model that incorporated a basic uncertainty process. This signal-detection model was described in the Appendix to that

article. The AAD for that model fit to the same data is only .03, a far closer fit than achieved by the associative-learning model. Not only that, but the cognitive model needs only three parameters to fit the data, not the five that Le Pelley uses. By any standard of parsimony in the modeling literature, the cognitive, three-parameter model that fits far better would be preferred.

Note also that the associative model's simulations come closest to fitting the data from Smith et al. (2006)—that is, the function of the fit index reaches its minimum—near Level 21 when the model predicts symmetry of the UR centered at Level 21 in alignment with the task's associative structure. Thus, the model fits best when it simply gives up predicting the crucial result that the macaque showed: an asymmetrical centering of the UR at Level 14, in sharp disalignment with the task's associative structure. It is a poor theoretical model that fits best when it simply gives up predicting the crucial result.

Associative processes are reactive to stimuli, shaped by reinforcement, entrained to payoff matrices and aligned to the task's

associative structure (Jozefowicz et al., 2009a; Le Pelley, 2012; Smith et al., 2008; Staddon et al., 2007). Every indication is that Murph's URs in Smith et al. (2006) were nonassociative. They occurred when the normal processes of conditioning were largely defeated. They occurred despite the UR's receiving no direct feedback. They occurred when the animal had no experience with trial-by-trial outcomes with which to form associations. And they occurred in a pattern showing that the animal's task construal had dissociated from the task's associative structure—probably because the task camouflaged that structure. An associative model of those URs qualitatively failed to predict their character. The URs in Smith et al. represent a notable nonassociative form of cognitive processing.

Nonetheless, these URs are easy to explain. Absent access to the task's associative structure, the animal constructed his own subjective decisional framework based on subjective difficulty, or uncertainty, or even metacognition. This subjective decisional framework (Figure 1C) centered near Trial Level 14 (not 21, the task's true breakpoint). This framework explains where his URs peaked and where he was indifferent to Sparse and Dense responses. It explains why he made about equal URs for Level 9 and Level 20 stimuli. He felt equally confident about those levels because they were equally far from his decisional breakpoint and equally clear to him as Sparses/Denses. Figure 1D shows generally that there was a strong relationship between the proportion of trials declined and the decisional distance of the trial level from the animal's decisional breakpoint ($r = .838$). Murph based his URs on a psychological signal that strongly correlated with subjective difficulty and uncertainty. Figure 1D shows that his use of the UR was quite precise. Therefore, one cannot just say that stimulus control was degraded by the deferred–rearranged feedback. To the contrary, Murph used the UR precisely for the trials that were subjectively most difficult for him.

Readers may benefit here from a perspective taken by a reviewer of this article. Murph clearly did not track the associative structure (Figure 1B)—his URs uncoupled from *objective* difficulty over a broad range of performance accuracies. But his URs were elegantly coupled to the task's *subjective* difficulty (Figure 1D) across the whole range of the task. Subjective difficulty is nonassociative because it is independent from reward, punishment, or reinforcement history. The organism can respond uncertain notwithstanding all of those factors. Murph did, suggesting that he was assessing difficulty subjectively and possibly metacognitively.

Uncertainty and Middle Responses (Beran, Smith, Coutinho, Couchman, & Boomer, 2009)

Synopsis of Beran, Smith, Coutinho, Couchman, and Boomer (2009)

Beran et al. (2009) asked whether capuchin monkeys (*Cebus apella*) could manage uncertainty in a Sparse–Dense paradigm. They gave six capuchin monkeys a version of the Sparse–Dense task that has already been described in this article. It presented difficult trials near the discrimination breakpoint that could be declined (with URs) to fend off errors. Strikingly, capuchin monkeys essentially did not use the UR (Figure 3A), even though in this experiment the capuchin monkeys received immediate feedback following every trial.

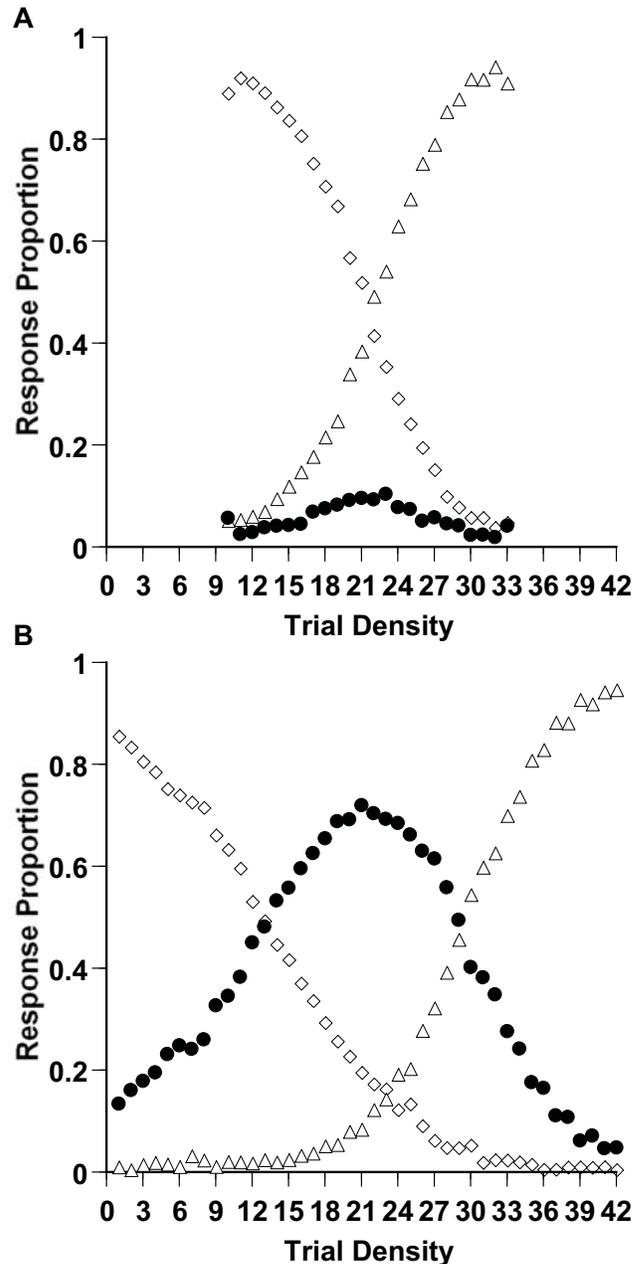


Figure 3. A. The performance of six capuchin monkeys in the Sparse–Uncertain–Dense task of Beran et al. (2009). The horizontal axis indicates the density level of the box. The proportions of trials ending with the Sparse response (open diamonds), Dense response (open triangles), and Uncertain response (filled circles) are shown for each trial level. B. The performance of the same capuchin monkeys in the Sparse–Middle–Dense task of Beran et al. (2009), depicted in a similar way. From *The Curious Incident of the Capuchins*, by J. D. Smith, M. J. Beran, J. J. Couchman, M. V. C. Coutinho, & J. B. Boomer, 2009, *Comparative Cognition and Behavior Reviews*, 4, 62.

To foster capuchin monkeys' URs, Beran et al. (2009) sharply increased the associative pressure toward it. They more than quadrupled the penalty timeout to 90 s from 20 s. Now, monkeys potentially gave up 30 trials during each error timeout (and poten-

tially 30 food rewards as well). This gave the capuchin monkeys a strong motivation to learn to make URs adaptively. But still, five of six monkeys did not do so.

The Sparse–Uncertain–Dense task for capuchin monkeys was interleaved with a Sparse–Middle–Dense task. In this task, the third response let the animals positively categorize middle stimuli as Middle, instead of making URs to acknowledge not knowing whether the stimuli were Sparse or Dense. Associative-learning mechanisms might well entrain MRs for middle stimuli in this three-response discrimination. And capuchin monkeys made MRs readily (Figure 3B).

Interpretation

Jozefowicz et al., (2009a, 2009b) did not address these findings from capuchin monkeys in their applications of an associative model to animal-metacognition data. Le Pelley (2012) also did not apply his associative-learning model to the capuchin monkeys' data patterns in either the Sparse–Uncertain–Dense case or the Sparse–Middle–Dense case. Yet these results are crucially important because they clearly undermine associative descriptions of URs.

The two tasks in Beran et al. (2009) were highly similar. Indeed, it was the same intermediate stimuli along the same perceptual continuum that should have recruited URs and MRs. The two tasks were mutual controls yet they created a sharp dissociation. The Sparse–Middle–Dense task let capuchin monkeys associate the MR to a class of middle discriminative stimuli. They accomplished this associative learning easily. The Sparse–Uncertain–Dense task, on the associative hypothesis, should have let capuchin monkeys associate URs to difficult discriminative stimuli. But the monkeys did not accomplish this associative learning easily at all. Capuchin monkeys are so responsive and intelligent that they are often called the poor-person's chimpanzee. They are not typically seen as associatively challenged. These animals would likely have shown the expected pattern of URs if a low-level associative strategy could produce it. To stress this point, if URs were a response to conflict, aversion, avoidance, fear, competing response strengths, reward maximization, hesitation-wavering behaviors, hesitation-wavering latencies, or any other potential associative cue, capuchin monkeys would have entrained to that cue and used it to occasion adaptive responding.

Yet they did not. Once again we are forced to conclude that there are two kinds of psychological signals and mechanisms at work in these tasks. The basis for MRs may be associative. The basis for URs is apparently not. Therefore, theory must explain psychologically why the mental states and representations that allow MRs but not URs are readily available to capuchin monkeys. This explanation will have to transcend describing performances vaguely and generally as associative because now we are considering the qualitatively different sorts of mental states and representations that capuchins can and cannot monitor.

Readers should also reflect on this evidence that the uncertainty-management systems of capuchin monkeys and macaques may be different. We could be seeing an aspect of reflective mind that is emerging more strongly in one branch of the primates than in another. Supporting this possibility, Basile, Hampton, Suomi, and Murray (2009) tested capuchin monkeys using a food-concealment paradigm designed by Call and Carpenter (2001). In this paradigm,

animals either see or do not see food being concealed in one of several tubes. Then, the question is whether they visually inspect the tubes to locate the food when they have not seen the hiding and so do not know its location, but gather in the food directly and immediately when they have seen the hiding and so do know its location. In one capuchin experiment, only one of five capuchins searched before reaching more often on trials where the hiding event was unseen compared with when it was seen. With eventual training, three of the five monkeys showed this pattern. But when the effort to search before responding was increased, no capuchin monkeys showed the uncertainty-monitoring pattern. Basile et al. concluded that the evidence for uncertainty monitoring in capuchins was equivocal. In a related study, Paukner, Anderson, and Fujita (2006) investigated capuchin monkeys' use of visual search to gain information. Food tubes could be visually searched before food selection. Some monkeys did use visual search on a limited basis to gain information before tube selection. However, they also searched transparent tubes unnecessarily (because the food was obviously visible), and searched bent tubes futilely (because the search could yield no pertinent information). Paukner et al. concluded that capuchin monkeys' behavior in their experiment contrasted with that of humans, apes, and macaques in similar studies. They suggested this could indicate species differences in metacognitive capabilities.

The capuchin-macaque dissociation—as it exists in the contemporary animal-metacognition literature—indicates that we should not just dismiss uncertainty-monitoring responses as associative. There is no associative chasm separating apes, macaques, and capuchins. But there may be a discontinuity in the higher-level cognitive mechanisms that let organisms cope with subjective difficulty and uncertainty. As the next section shows, researchers are working toward understanding these higher-level processes.

Executive Uncertainty Responses (Smith, Coutinho, Church, & Beran, 2013)

Synopsis of Smith, Coutinho, Church, and Beran (2013)

Smith et al. (2013) explored these higher-level processes. To do so, they added a secondary task requirement to the ongoing discrimination performance of macaques. The secondary task acted as a concurrent cognitive load, and it was possible to then examine the effect of the load on the use of the various responses in the task.

Four rhesus macaques performed a Sparse–Uncertain–Dense task or a Sparse–Middle–Dense task like those already described. The UR let them decline difficult trials near their discrimination thresholds as well as any other trials of their choosing. The MR let them perceptually classify middle stimuli. A requirement to perform concurrent memory tasks was occasionally added to ongoing perceptual-discrimination performance.

The experiments considered how the concurrent memory load might affect perceptual-classification processes as compared with the processes that support uncertainty monitoring and URs. Smith et al. (2013) suggested that perceptual classification could be more stimulus-based, reactive, and associative, making few demands on working memory or executive attention. In that case, Sparse, Dense, and possibly Middle responses might be minimally affected

by the working memory load. In contrast, Smith et al. suggested that URs might be strongly affected by the concurrent load—if they especially depend on executive functioning or attentional resources.

Figure 4 shows the article's main results. Concurrent tasks disrupted macaques' URs far more than their Sparse, Middle, or Dense perceptual responses. In particular, MRs survived the concurrent load nearly intact. For one macaque, they survived completely intact.

Interpretation

It is not a criticism of associative models in the animal-metacognition literature to say that so far none of them has engaged the phenomenon in Smith et al. (2013). There has not

been time to address it yet. Still, associative models have no way to explain the dissociation produced by a concurrent cognitive load. The differential effects of concurrent loads on URs and MRs are important because these responses were associated with the same stimuli along the same stimulus continuum. One cannot argue that the effect on URs arose just because intermediate stimuli are associatively challenging or because difficult perception is resource intensive. Perceptual difficulty can only be part of the story because MRs were less disrupted by the memory load. The dissociation shown by capuchin monkeys between URs and MRs (Beran et al., 2009) also showed that perceptual difficulty is not the whole story. Indeed, one cannot pursue any model that portrays the UR as an associative reaction to stimuli just as the MR is. Any such model fails qualitatively because the two responses

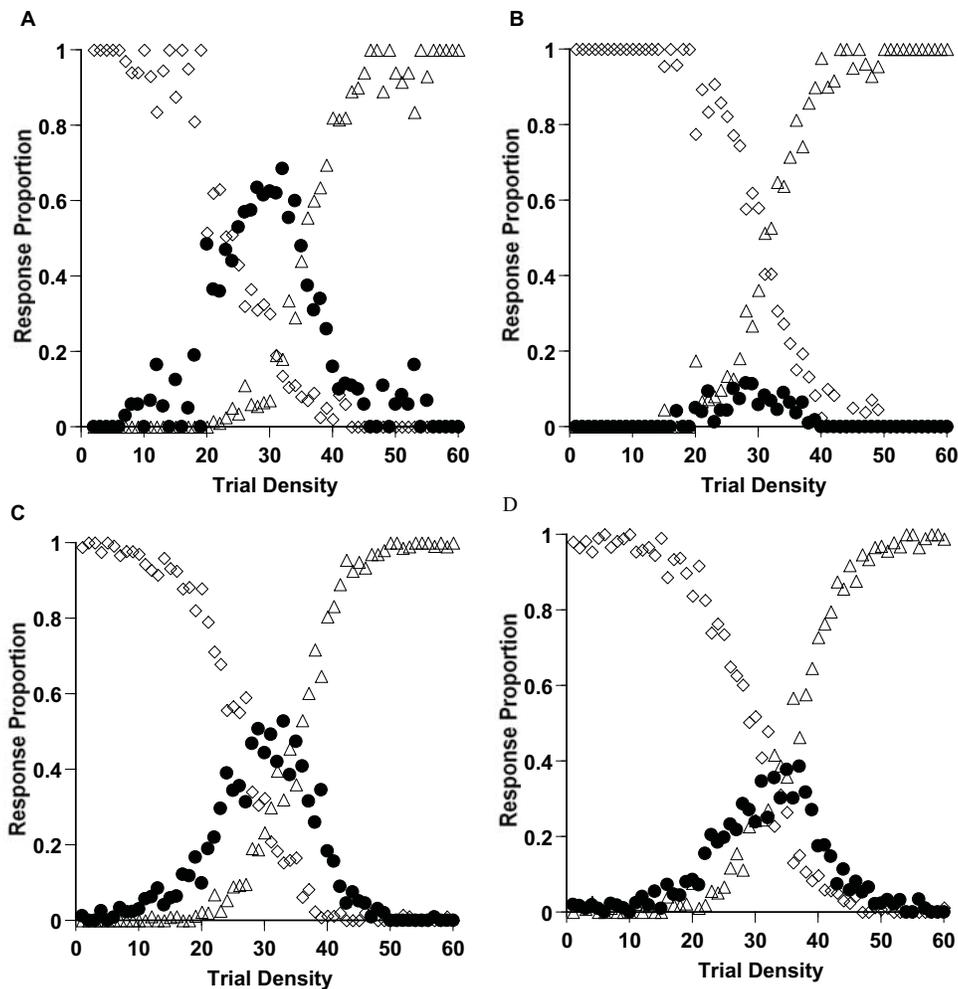


Figure 4. A, B. Proportion of uncertainty responses (solid circles), sparse responses (open diamonds), and dense responses (open triangles) made by macaques Murph and Lou in their baseline performance and in their first phase of concurrent-load testing. C, D. Percentage of middle responses (solid circles), sparse responses (open diamonds), and dense responses (open triangles) made by macaques Hank and Gale in their baseline performance and in their first phase of concurrent-load testing. From *Executive-Attentional Uncertainty Responses by Rhesus Macaques (Macaca mulatta)* by J. D. Smith, M. V. C. Coutinho, B. A. Church, & M. J. Beran, 2013, *Journal of Experimental Psychology: General*. Copyright, 2013 by the American Psychological Association. Reprinted with permission.

behave qualitatively differently in the face of cognitive resources occupied.

The problem here for associative accounts is that they assume a theoretical equivalence among the various responses in uncertainty-monitoring tasks. They make no allowance for the possibility that these responses might have qualitatively different requirements in resources and commitments at different levels within the cognitive system. But they evidently do have different requirements and different cognitive levels. Therefore, any artificial label that lumps them together—including the label associative—is misleading and theoretically inappropriate. Such a label blurs important theoretical distinctions.

In contrast, [Smith et al. \(2013\)](#) provided an information-processing description of URs that specifies their nonassociative character. Appropriate use of the UR is more demanding of working memory resources. It is more executive and attentional in psychological character. This dissociation suggests that the uncertainty response is a higher-level, decisional response that is particularly dependent on working memory and attentional resources. This is consistent with the theoretical possibility that the UR is an elemental behavioral index of uncertainty monitoring or metacognition.

The result in [Smith et al. \(2013\)](#) mirrors human research in which participants performed primary memory tasks while reporting metacognitive tip-of-the-tongue states and judgments of learning ([Schwartz, 2008](#)). Schwartz found that concurrent working memory loads strongly affected both kinds of metacognitive judgment, in particular sharply decreasing reports of tip-of-the-tongue experiences. Schwartz concluded that working memory and metamemory use similar monitoring processes, a conclusion analogous to that made here. The present result also mirrors human research by [Paul, Smith, and Ashby \(2014\)](#) that used rapid event-related functional MRI to explore the neural systems underlying humans' uncertainty responses as distinct from their primary perceptual responses. These scanning results indicated that the uncertainty response is not just a middle category but rather that uncertainty monitoring is mediated by a distinct network that includes anterior cingulate cortex, prefrontal cortical areas, and insula. Neuroimaging could make an important contribution to the comparative literature on this topic, to the extent that these studies with nonhuman primates eventually become feasible.

Memory Monitoring by Macaques ([Hampton, 2001](#))

Synopsis of [Hampton \(2001\)](#)

The results in [Smith et al. \(2013\)](#) also offer a new perspective on [Hampton's \(2001\)](#) metamemory studies. In these experiments, macaques performed a delayed matching-to-sample task. The UR let monkeys accept a less preferred but guaranteed food reward instead of attempting the memory test to possibly gain a preferred reward. Sometimes macaques did not have the UR option available and were forced to complete the memory test. Other times, they could make a free choice between declining or completing the memory test.

Hampton reported three principal results ([Figure 5A](#)). First, not surprisingly, with longer delays between sample presenta-

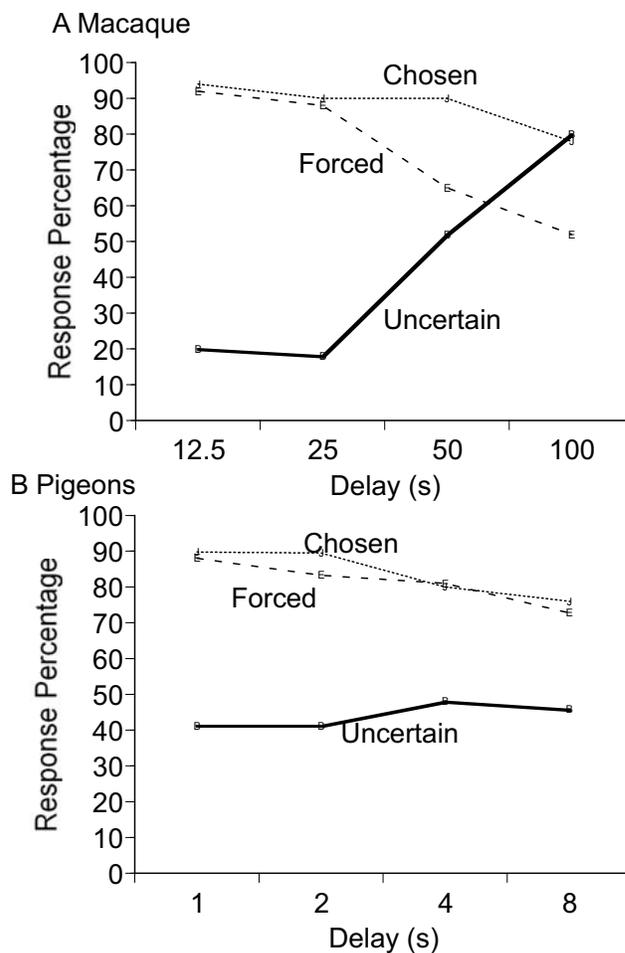


Figure 5. A. Memory performance by a macaque in the delayed matching-to-sample task of [Hampton \(2001\)](#). The horizontal axis indicates the length of the retention interval before matching could occur. The percentage of trials that received the uncertainty response is shown (solid line). The percentages correct of memory tests completed are also shown, on occasions when the memory test was mandatory (dashed line) or optional and voluntarily selected by the macaque (dotted line). B. Memory performance by a pigeon in the delayed matching-to-sample task of [Inman and Shettleworth \(1999\)](#).

tion and match-choice selection, matching performance decreased because macaques remembered less well which picture they had seen. Their memories were fading and failing. Second, macaques selectively declined memory tests at long retention intervals—appropriately so because they had mostly forgotten the sample and would remember less correctly if they attempted the memory test. Third, the macaques were able to sustain strong performance even at long retention intervals *when they got to choose when they would complete the memory test*. This suggests that they were monitoring some felicitous internal signal of remembering that predicted well for successful matching. [Templer and Hampton \(2012\)](#) presented additional experiments that reinforced this conclusion.

Interpretation

These macaques cannot have been conditioned to avoid particular stimuli. Hampton used sets of four stimuli in each day of testing. The same stimuli were the foils and targets on each trial, and were each equally likely to be the correct choice on any given trial. Thus, long-term avoidance learning about particular stimuli was never useful. Moreover, the UR was made when there were no sample or foil stimuli available that could trigger approach or avoidance responses. Nor is it that macaques just learned that long retention intervals meant scant rewards, so that those trials were avoided through URs. To evaluate this associative response to the length of the retention interval, Hampton compared URs on memory trials (when a memory sample was presented) to blank trials (when no memory sample was presented and there was nothing to remember). Now, with delay interval controlled, animals made far more URs on blank trials. Macaques were responding to the monitored strength of available memory traces and not just responding associatively to the length of the retention interval. Thus, Hampton's macaques may have shown a form of metamemory (see Kornell et al., 2007, for a similar result). They monitored memory's contents to decline tests of weaker memories. This memory-strength signal—abstract and cognitive—would seem to be profoundly different from the stimulus signal available in traditional associative-learning situations, and profoundly underserving of the label associative.

Nonetheless, Le Pelley (2012) tried to fit an associative-learning model to these results. He assumed a register of trace-strength activation levels corresponding to the stimuli that could be presented in the task. Simulated performers would associatively accept trials when they encountered a strong trace strength available in the register. Trials would be associatively avoided (with the UR when that response was available) when they encountered no sufficiently strong trace strength available in the register.

If this arrangement still sounds like memory traces are being monitored, and trials accepted if the sample stimuli were remembered well, then one understands the model correctly. We will return to this conceptual issue but let us take the model on its own terms for a moment.

It is presently difficult to judge the model's pragmatic success. In implementing the model, Le Pelley (2012) set the reward value for a UR in Hampton's task to be 0. However, Hampton always positively rewarded URs with a food pellet. We contacted Hampton to confirm this. Therefore, the UR would have to be given some positive reinforcement value in the model. But, if one does so straightforwardly, the model qualitatively fails to recover Hampton's data. Le Pelley confirmed these points but notes that the model can still be given parameter settings that reproduce Hampton's data. We accept that judgment. Nonetheless, the published simulations must be discounted because they misstate a crucial aspect of Hampton's methodology. As presented, they do not support an associative hypothesis about the memory-monitoring results or contribute to the study of animal metacognition.

In another associative approach, Jozefowicz et al. (2009a) showed that an associative model similar to Le Pelley's produced a data pattern that was qualitatively unlike Hampton's data. They could only recover Hampton's data if they also assumed that animals were responding associatively to the length of the reten-

tion interval. But remember that Hampton with his blank trials discounted this temporal cue. Therefore, the associative model of Jozefowicz et al. also cannot be judged entirely successful.

There are additional memory-monitoring results from animals that neither associative model can explain. For example, in Smith, Shields, Allendoerfer, and Washburn (1998), macaques saw a list of sample pictures and then judged whether a probe picture had been in that list. Macaques could make URs to decline any memory tests they chose. Macaques remembered early and late list items best when they accepted memory tests (Figure 6A). They made URs in the mirror image of that performance pattern, declining memory tests most when difficult list positions were tested. In addition, macaques held their error rate constant at 10% as the memory lists grew longer and the memory tests harder (Figure 6B). They accomplished this by making more URs for longer/harder lists (Figure 6C).

The associative models would produce about equal trace strengths for all the samples shown in a memory list and, thus, equal rates of accuracy and URs across the serial positions. They cannot recover the primacy and recency effects that animals showed. Now one could add to the models a primacy parameter, so that list-initial items entered the system stronger. One could add a recency parameter, so that list-final items were more available. Notice how the associative system is quietly starting to look like a memory system—here is the problem again that was mentioned in passing in discussing Le Pelley's memory simulation. Notice also that on adding these parameters the associative model in Le Pelley's case would have nine free parameters (i.e., response variation, memory variation for foils, memory variation for samples, the primacy parameter, the recency parameter, reward value for errors, reward values for corrects, reward values for bailouts, and a scaling parameter). In long experience with formal modeling in comparative and cognitive psychology (e.g., Smith, 2002, 2005, 2006; Smith & Minda, 2001; Smith et al., 2008), we have never seen a model with this number of free parameters that was deemed plausible. To be clear, neither associative modeler proposed this nine-parameter simulation, though something like it would be required to account for the results in Smith et al. (1998). This would be an inappropriate level of mathematical flexibility to bring to any psychological phenomenon and perhaps this is why the associative modelers wisely did not make the attempt.

In contrast, everything in Smith et al. (1998)—the primacy, the recency, the constant error rate, the increasing bailout rates with increasing difficulty—is intuitively explained if animals are only granted a basic capacity to monitor their memories. Then, they opt out of trials when the traces they monitor are neither strong enough to report memory, or weak enough to conclude that the sample must not have been shown in the previous list of pictures.

In fact, Smith, Shields, and Washburn (2003) fit formal models to monkeys' performance in both Smith et al. (1998) and Hampton (2001), using an intuitive memory-monitoring model. They found that monkeys defended the same criterion of memory strength in both studies, a striking convergence across laboratories, animals, and methodologies. It is an important additional consideration that a memory-monitoring interpretation easily and naturally bridges across the experiments in Hampton and Smith et al., whereas the associative models cannot. Thus, the associative models in the animal-metacognition literature also fail the test of theoretical parsimony.

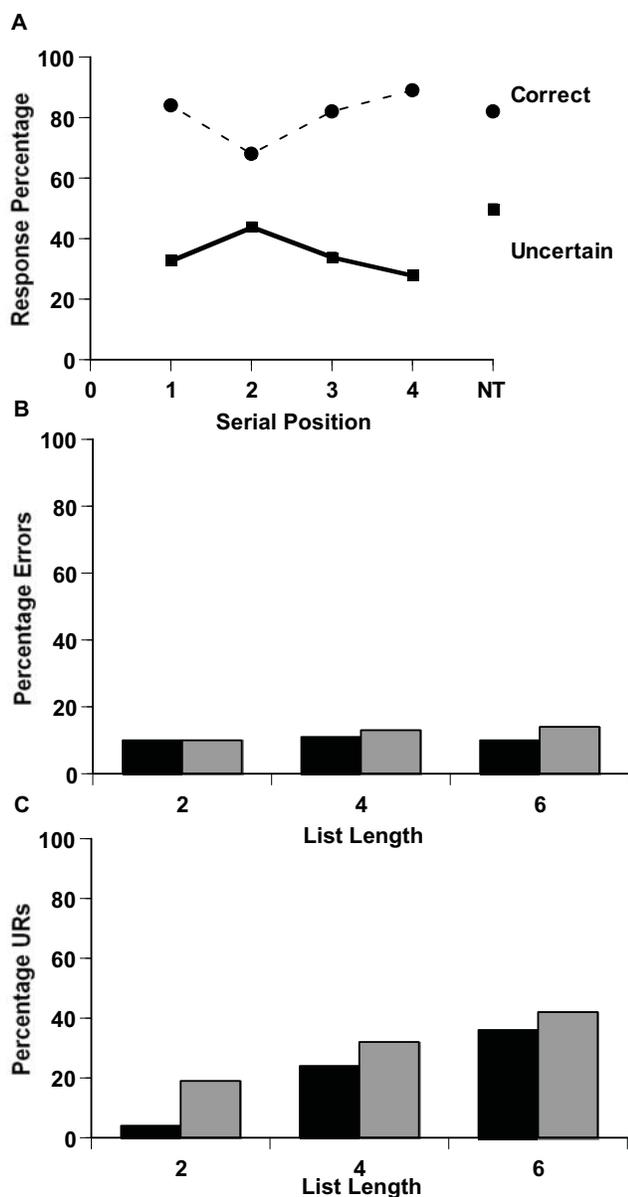


Figure 6. A. Memory performance by a macaque in the metamemory task of Smith et al. (1998). NT denotes probe pictures that were “Not There” in the memory list of pictures. The serial position (1–4) of the probe picture in the list of pictures is also given along the X-axis for probes on “There” trials. The percentage of total trials that received the uncertainty response is shown (solid line). The percent correct (of trials on which the memory test was accepted) is also shown (dashed line). B. Percentage error rates by two macaques (black and gray bars) when the difficulty of the memory test was increased by increasing the length of the memory list from two, to four, to six pictures. C. Percentage uncertainty responses (URs) by the two macaques when the difficulty of the memory test was increased in the same way. From *Memory Monitoring by Humans and Animals*, by J. D. Smith, W. E. Shields, K. R. Allendoerfer, & D. A. Washburn, 1998, *Journal of Experimental Psychology: General*, 127, 236, 238. Copyright, 1998 by the American Psychological Association. Reprinted with permission.

Misunderstanding Associative Learning

Ultimately, these associative models raise even more basic concerns. These models embody an understanding of associative learning that in our view is incorrect and harmful to theoretical development within comparative psychology. We discuss this issue now, an issue that no amount of remedial simulation (regarding Hampton et al.’s studies) or additional simulation (regarding Smith et al.’s studies) could address.

The memory simulations are premised on the idea that animals can learn to respond on the basis of the internal, representational activation of a stimulus. Documenting this ability, Le Pelley (2012) cites articles by Mishkin and Delacour (1975) and Overman and Doty (1980). These articles explored this capacity in macaques using the well-known Delayed-Matching-to-Sample paradigm, in which animals see a sample picture, and then, after a retention interval, see a set of choice stimuli. They must select the more familiar or recently seen stimulus in a matching paradigm (i.e., they must match the original sample).

In these tasks, there are stimuli on the scene at the moment of behavioral choice that are memory or trace-strength relevant. As the animal looks at the test stimuli, those stimuli may automatically cue the animal as to their internal trace strength or hotness, and this hotness might become a cue underlying associative responding in a task.

However, Hampton’s methodology is strikingly different. His macaques, after observing the sample and waiting through the retention interval, are presented with two abstract response icons that have no trace-strength relevance. The macaque must choose the response icon that either accepts or cancels the memory test. In this case, there are no memory-relevant stimuli on the scene. There is nothing to cue the animal’s response to some stimulus’s trace hotness. What the animal must do is use those response icons to cue a memory search for a hot or recent trace. It is true to say that the search needs only consider the four stimuli that were recycled during each day of testing but it is a search of memory locations nonetheless. Moreover, the hot trace to be searched for cannot be the now-present response icons, or the animal’s monitor, or the animal’s cage, or the animal’s labmate, or the animal’s juice dispenser, or anything like that. The animal must restrict the search for trace strength to the items that the experiment is generally testing for. Nothing else can help. And nothing on the response screen can cue the macaque associatively to conduct this search of memory or to restrict that search to the required set of memory locations, or on how to judge when a remembered item has been retrieved or has not been retrieved.

Thus, analyzing the information-processing requirements of Hampton’s task carefully, one realizes that one has left associative learning and responding qualitatively behind. These concepts might apply to Mishkin and Delacour (1975) but they do not apply to Hampton (2001). For the same reason, Jozefowicz et al. (2009a) are incorrect when they claim that the underlying processes are exactly the same whether a triggering stimulus is on the screen or the animal must initiate its own targeted search for active memory traces. The triggering stimulus can be reacted to associatively. Facing the memory-neutral icons, the macaque must have a task construal, conduct the memory search, restrict it appropriately, evaluate it adaptively, and so forth. The concurrent memory-load results of Smith et al. (2013) further suggest that the processes that

support URs in monitoring tasks are executive and attentional, not automatic and reactive. Moreover, Metcalfe (2008) made stimulus absence in metacognitive tasks a crucial benchmark in her operational definition of animal metacognition. Therefore, viewed from all these perspectives, the performances observed in Hampton (2001) cannot properly be considered to be based on associative learning, and overlooking this point is a significant error.

The theoretical problem is that the associative modelers in this area have fallen into a formal-mathematical trap of which all experimental psychologists must be aware. Their simulations all include internal, representational registers indexed as a series of perceptual or memorial psychological states. These registers contain abstract activation strengths so that some positions in the register can produce one response and other positions produce another response. The registers are updated based on the rewards and penalties for correct and incorrect responses so that the register's activations reflect the reinforcement history in the task.

The trap is that this recurring structure is abstract, formal, and mathematical. It does not attend to the real task that the organism faces, or to the information-processing requirements of the task. Indeed, forcing all performances onto this structure artificially equates them—really, this is what you want to do and what you have to do if you want to label them all associative. Unfortunately, then the associative label becomes meaningless, because one can no longer make distinctions among performances that may be highly distinctive theoretically and psychologically.

To illustrate this idea, we took one more step along the path of modeling performance in an uncertainty-monitoring task. We imagined that animals have an internal register of fully aware, conscious, declarative, episodic, metacognitive states—states of remembering or not remembering. These would run along the register from very weak, dimly remembered items to floridly, episodically remembered items. Of course the animal would gain rewards when it accepted memory tests of consciously blazing memories. Of course it would benefit by bailing out when it consciously knew it did not know. Through these outcomes, the conscious-metacognitive register would be updated by reinforcement just as associative-learning models assume. It would pay to answer memory trials when you consciously know you remember. It would pay to bail when you consciously know you do not. In this way, we could claim mathematically and formally that animals learn associatively through reinforcement to use fully conscious metacognitive states.

This claim is useless. The animals would still be consciously metacognitive, no matter whether reinforcement was involved and no matter whether one calls it associative. Moreover, this claim implies that any kind of information processing can be lumped under the construct associative, which makes that term theoretically useless. In short, this thought experiment demonstrates why the registers, their activations, their updatings by reinforcement, are psychologically empty, abstract, and meaningless—they cannot bear on the psychological representations and processes that the organism brings to the task, or on the psychological level they have in the animal's cognitive system. And then we have to consider carefully which the comparative psychologist should care more about: the abstract and meaningless structures and registers or the true psychology of the situation.

To us, this answer is clear. The psychology matters. The nature of the information processing matters. The psychological content

of the registers matters. The way the registers are accessed matters. The executive processes required and the attentional effort demanded matter. The animal's awareness level of the registers matters. These conclusions mesh well with those of Jozefowicz et al. (2009a, 2009b), who emphasize that the reinforcement contingencies in these situations do not matter—rather, it is the cognitive content of the processes and operations that matter. We believe these conclusions also mesh with the views of other comparative psychologists working today.

Too often, associative-learning theorists take the surface structure of a situation to reveal associative processing writ abstractly, neglecting to analyze carefully the real cognitive content and information-processing stages that an animal's task requires. The associative models in the animal-metacognition literature exemplify this serious error.

The Phylogenetic Map of Animal Metacognition

By understanding this error, one gains a new perspective on the animal-metacognition literature. We will illustrate this perspective using the research that has tried to map the distribution of metacognition more broadly across the vertebrates, especially beyond the primates.

Memory Monitoring by Pigeons

For example, several studies on pigeons (Teller, 1989; Inman & Shettleworth, 1999; Sutton & Shettleworth, 2008) have explored pigeons' memory-monitoring capacity using paradigms that are similar to those used by Hampton (2001). The general pattern of results for macaques and pigeons are contrasted in Figures 5A and 5B, respectively. Simply put, pigeons show no components of the uncertainty-monitoring capacity shown by Hampton's macaques. Their UR curve stays flat over the range of retention intervals—they do not make more URs when given longer to forget (though they should because they remember less after longer delays). They do not perform better on the memory tests they choose to complete compared with the trials they are forced to complete, as they would if they were monitoring some felicitous signal of memory strength or subjective remembering.

Given this negative result in Inman and Shettleworth (1999), Sutton and Shettleworth (2008) added multiple experiments. They confirmed that pigeons do not express any memory-monitoring capacity. In short, the pigeon results from Teller (1989), until the present, are different from the metamemory and confidence-rating results produced by macaques in Hampton (2001), Kornell et al. (2007), Shields, Smith, Guttmannova, and Washburn (2005), and Smith et al. (1998), though efforts continue to confirm robust memory-monitoring processes in pigeons (e.g., Adams & Santi, 2011). An associative hypothesis has difficulty explaining this species difference. Pigeons are adept associative learners. They clearly have stimulus registers updated by reinforcement. If there were low-level cues in these tasks that supported associative learning, it is likely they would find them. Clearly, it is unhelpful in this case to advance the vague surface-structure definition of "associative." That is, one cannot say: Well, it is all associative, but metacognitive tasks involve a kind of associative learning that macaques but not pigeons can do.

In fact, the pigeon-macaque dissociation, like the capuchin-macaque dissociation, like the concurrent-load dissociation, shows

that there are important differences across different performances. Therefore, theory must explain psychologically why the mental states and representations that allow efficient performance in the memory-monitoring tasks are available to macaques but not to pigeons. Theory must engage the deep structure of the animal's cognitive performance. That engagement lies outside of associative-learning theory because then we are considering different sorts of mental states and processes—like the processes of monitoring memory-trace strength—that pigeons cannot easily access but macaques can.

Information Seeking by Pigeons

As another example, Roberts et al. (2009) studied information-seeking responses—another possible reflection of metacognition—by pigeons in a matching-to-sample task. Roberts et al. asked simply whether pigeons would realize that they needed to make a response to reveal an occluded sample before trying to choose the shape that matched the sample. In several experiments, pigeons did not make this response that let them see or study the sample. They just kept choosing the match choices, trying to match the sample that of course they had not seen and could not match. Beran and Smith (2011) gave the Roberts et al. test and extensions of it to macaques. Macaques understood that they had to reveal the sample if they hoped to match it successfully. They also flexibly dealt with other variations, such as having a sample but not the match choices, or having available all necessary information to answer correctly. They requested only the information they were missing and they simply matched when they had the information they needed.

From an associative perspective, if the response to uncover the sample before trying to match wins rewards, avoids punishments, and rescues the animals from aversive environments, then why don't both of these associatively adept species do that? Why don't pigeons form an associative connection to the reveal-the-sample response? It is highly rewarding when they make it. From an associative perspective, the answer is unclear (but see Zentall & Stagner, 2010). From an uncertainty-monitoring perspective, though, the answer is intuitive. Macaques know they lack the relevant information to match successfully; pigeons do not. Of course one need not attribute full, conscious, self-imbued metacognition to macaques even if one adopts this information-seeking hypothesis. Nonetheless, this hypothesis describes the species difference when the associative hypothesis does not. In short, pigeons, by barely expressing uncertainty-monitoring capacities, underscore the cognitive complexity, the distinctive psychology, and the nonassociative character of uncertainty responding.

Summarizing the current state of the animal-metacognition literature, Carruthers and Ritchie (2012) concluded: "We fully accept that this body of work, taken as a whole, cannot be explained in low-level associationist terms, as involving mere conditioned responses to stimuli. A great deal of careful experimentation has been done to demonstrate that this is not the case, and we are happy to embrace this conclusion (Beran et al., 2009; Couchman et al., 2010; Smith, Redford, Beran, & Washburn, 2010; Washburn, Gullege, Beran, & Smith, 2010)." Accordingly, readers will want to fit in their own way the associative models discussed in this article into the larger context of the field's present understanding,

and to weigh the perspective of the associative modelers against the perspectives of others.

Conclusion

The associative models in the animal-metacognition literature offer an illuminating case study in what comparative psychology has sometimes been over its 100 years. Those models reflect an ultimately strict adherence to Morgan's (1906) Canon, which even Morgan did not intend and which many comparative-cognition researchers do not accept now. They offer a universalist definition of associative learning (the registers, the reinforcements, the updates), and a universalist null hypothesis claiming that animals' uncertainty performances can be explained in purely associative ways. They insist on a qualitative distinction between metacognitive and associative performances, when, of course, there could be levels and gradations between, involving executive function, attentional resources, and so forth as precursors to full awareness and declarative cognition. Even in Le Pelley's (2012) title—*Metacognitive Monkeys or Associative Animals . . .*—he seems to recommend a yes or no answer to the question of animal metacognition, when that dichotomy could be wrong for the field's theoretical goal.

In our view, it is important to reflect on this vision for comparative psychology, and to see the harm it potentially does. This vision exaggerates the conceptual distance between human and animal minds. It ignores that phylogeny probably gave reflective mind and metacognition to animals in stages of attainment, so that the stages of the emergence from monkeys to apes to humans could be crucial in understanding the rise of our reflective mind. It minimizes the potential value of animal models for metacognition. These models—based on monkeys' purely associative processes—could have no useful role in studying humans' truly metacognitive processes. Thus, this vision downgrades the relevance of animal research to human researchers. It cuts cognitive and comparative researchers off unnecessarily from constructive interactions. Historically, it even led us to underestimate the mental experience of animals, including their pain and suffering.

In contrast, recent developments in animal-metacognition research illustrate an alternative approach to comparative research and theory. The research clearly suggests that what animals do in some tasks is higher-level and cognitive, deliberate and decisional, and possibly metacognitive and conscious. Therefore, this area of research joins others to show that comparative psychology can sometimes be about animal-human continuities. It can avoid jumbling together—under the rubric "associative"—animal performances that are importantly different in psychological character. It can attempt to understand the deep structure of animals' cognitive performances, in the abstractness of their representations, in the complexity of their processes, and in their attentional/executive psychological organization. It can acknowledge that a capacity like metacognition may not be all or none, and that animals might have a basic uncertainty-monitoring system without having bought all the apps of human metacognition, like consciousness, self-awareness, and even self-loathing from cognitive failure and indecision.

In our view, this alternative approach to theoretical interpretation has many strengths. This approach could lead us to ask about the earliest and preverbal developmental roots of human metacognition (Balcomb & Gerken, 2008). Moreover, the animal paradigms dramat-

ically expand the range of paradigms available for testing young humans. This approach could suggest that there are basic, nonverbal forms of cognitive regulation that could be fostered in children—for example, in language-delayed and autistic children, or children with mental retardation—who can't reach the explicit, declarative peaks of cognitive self-awareness. This approach suggests, therefore, that animal models of metacognition do have significance, and that animal-metacognition research has intrinsic relevance to issues of education, training, and humans' self-regulation and behavioral control. This approach makes it clear why human cognitive psychologists should be aware of and sharply interested in the growing and influential study of uncertainty processes in animals. And, thus, comparative science in this domain is constructively woven into the fabric of contemporary cognitive science.

References

- Adams, A., & Santi, A. (2011). Pigeons exhibit higher accuracy for chosen memory tests than for forced memory tests in duration matching-to-sample. *Learning & Behavior, 39*, 1–11.
- Balcomb, F. K., & Gerken, L. (2008). Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science, 11*, 750–760. doi:10.1111/j.1467-7687.2008.00725.x
- Basile, B. M., Hampton, R. R., Suomi, S. J., & Murray, E. A. (2009). An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 12*, 169–180. doi:10.1007/s10071-008-0180-1
- Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General, 127*, 55–68. doi:10.1037/0096-3445.127.1.55
- Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Cognition, 120*, 90–105. doi:10.1016/j.cognition.2011.02.016
- Beran, M. J., Smith, J. D., Coutinho, M. V. C., Couchman, J. J., & Boomer, J. (2009). The psychological organization of “uncertainty” responses and “middle” responses: A dissociation in capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes, 35*, 371–381. doi:10.1037/a0014626
- Beran, M. J., Smith, J. D., & Perdue, B. M. (2013). Language-trained chimpanzees (*Pan troglodytes*) name what they have seen but look first at what they have not seen. *Psychological Science, 24*, 660–666. doi:10.1177/0956797612458936
- Call, J. (2010). Do apes know that they can be wrong? *Animal Cognition, 13*, 689–700. doi:10.1007/s10071-010-0317-x
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 3*, 207–220. doi:10.1007/s100710100078
- Carruthers, P. (2008). Meta-cognition in Animals: A skeptical Look. *Mind & Language, 23*, 1, 58–89. doi:10.1111/j.1468-0017.2007.00329.x
- Carruthers, P., & Ritchie, J. B. (2012). Metacognition and uncertainty in animals. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *The foundations of metacognition* (pp. 76–93). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199646739.003.0006
- Couchman, J. J., Coutinho, M. V. C., Beran, M. J., & Smith, J. D. (2010). Beyond stimulus cues and reinforcement signals: A new approach to animal metacognition. *Journal of Comparative Psychology, 124*, 356–368. doi:10.1037/a0020129
- Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *Handbook of memory and metamemory*. New York, NY: Psychology Press.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911. doi:10.1037/0003-066X.34.10.906
- Foote, A. L., & Crystal, J. (2007). Metacognition in the rat. *Current Biology, 17*, 551–555. doi:10.1016/j.cub.2007.01.061
- Fujita, K. (2009). Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 12*, 575–585. doi:10.1007/s10071-009-0217-0
- Gallup, G. G. (1982). Self-awareness and the emergence of mind in primates. *American Journal of Primatology, 2*, 237–248. doi:10.1002/ajp.1350020302
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences of the United States of America, 98*, 5359–5362. doi:10.1073/pnas.071600998
- Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews, 4*, 17–28.
- Inman, A., & Shettleworth, S. J. (1999). Detecting metamemory in non-verbal subjects: A test with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes, 25*, 389–395. doi:10.1037/0097-7403.25.3.389
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. (2009a). Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews, 4*, 29–39.
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. (2009b). Reinforcement and metacognition. *Comparative Cognition and Behavior Reviews, 4*, 58–60.
- Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*, 609–639. doi:10.1037/0033-295X.100.4.609
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–326). Cambridge, UK: Cambridge University Press. doi:10.1017/CBO9780511816789.012
- Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Cognitive Science, 18*, 11–15. doi:10.1111/j.1467-8721.2009.01597.x
- Kornell, N., Son, L., & Terrace, H. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*, 64–71. doi:10.1111/j.1467-9280.2007.01850.x
- Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*, 686–708. doi:10.1037/a0026478
- Metcalf, J. (2008). Evolution of metacognition. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of memory and metamemory* (pp. 29–46). New York, NY: Psychology Press.
- Metcalf, J., & Kober, H. (2005). Self-reflective consciousness and the projectable self. In H. S. Terrace & J. Metcalfe (Eds.), *The missing link in cognition: Origins of self-reflective consciousness* (pp. 57–83). New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195161564.003.0002
- Metcalf, J., & Shimamura, A. (1994). *Metacognition: Knowing about knowing*. Cambridge, MA: Bradford Books.
- Mishkin, M., & Delacour, J. (1975). An analysis of short-term visual memory in the monkey. *Journal of Experimental Psychology: Animal Behavior Processes, 1*, 326–334. doi:10.1037/0097-7403.1.4.326
- Morgan, C. L. (1906). *An introduction to comparative psychology*. London, UK: Walter Scott.
- Nelson, T. O. (Ed.). (1992). *Metacognition: Core readings*. Toronto, Canada: Allyn & Bacon.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116. doi:10.1037/0003-066X.51.2.102
- Overman, W. H., & Doty, R. W. (1980). Prolonged visual memory in macaques and man. *Neuroscience, 5*, 1825–1831. doi:10.1016/0306-4522(80)90032-9
- Paukner, A., Anderson, J. R., & Fujita, K. (2006). Redundant food searches by capuchin monkeys (*Cebus apella*): A failure of metacognition? *Animal Cognition, 9*, 110–117. doi:10.1007/s10071-005-0007-2

- Paul, E. J., Smith, J. D., & Ashby, F. G. (2014). *Neural correlates of uncertainty monitoring in human categorization*. Manuscript submitted for publication.
- Roberts, W. A., Feeney, M. C., McMillan, N., MacPherson, K., & Musolino, E. (2009). Do pigeons (*Columba livia*) study for a test? *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 129–142. doi:10.1037/a0013722
- Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review*, *1*, 357–375. doi:10.3758/BF03213977
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition*, *36*, 9–19. doi:10.3758/MC.36.1.9
- Serra, M. J., & Dunlosky, J. (2005). Does retrieval fluency contribute to the underconfidence-with-practice effect? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 1258–1266. doi:10.1037/0278-7393.31.6.1258
- Shields, W. E., Smith, J. D., Guttmanova, K., & Washburn, D. A. (2005). Confidence judgments by humans and rhesus monkeys. *Journal of General Psychology*, *132*, 165–186.
- 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*, 147–164. doi:10.1037/0096-3445.126.2.147
- Smith, J. D. (2002). Exemplar theory's predicted typicality gradient can be tested and disconfirmed. *Psychological Science*, *13*, 437–442. doi:10.1111/1467-9280.00477
- Smith, J. D. (2005). Wanted: A new psychology of exemplars. *Canadian Journal of Experimental Psychology*, *59*, 47–53.
- Smith, J. D. (2006). When parameters collide: A warning about cognitive models. *Psychonomic Bulletin & Review*, *13*, 743–751. doi:10.3758/BF03193991
- Smith, J. D., Beran, M. J., & Couchman, J. J. (2012). Animal metacognition. In T. Zentall & E. Wasserman (Eds.), *Comparative cognition: Experimental explorations of animal intelligence* (pp. 282–304). Oxford, UK: Oxford University Press.
- Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C., & Boomer, J. (2009a). Animal metacognition: Problems and prospects. *Comparative Cognition and Behavior Reviews*, *4*, 33–46.
- Smith, J. D., Beran, M. J., Couchman, J. J., Coutinho, M. V. C., & Boomer, J. (2009b). The curious incident of the capuchins. Problems and prospects. *Comparative Cognition and Behavior Reviews*, *4*, 47–50.
- Smith, J. D., Beran, M. J., Coutinho, M. V. C., & Couchman, J. J. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin & Review*, *15*, 679–691. doi:10.3758/PBR.15.4.679
- Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty states and reinforcement signals in the comparative study of metacognition. *Journal of Experimental Psychology: General*, *135*, 282–297. doi:10.1037/0096-3445.135.2.282
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2012). The highs and lows of theoretical interpretation in animal metacognition research. *Philosophical Transactions of the Royal Society B*, *367*, 1297–1309. doi:10.1098/rstb.2011.0366
- Smith, J. D., Coutinho, M. V. C., Boomer, J., & Beran, M. J. (2012). Metacognition across species. In J. Vonk & T. K. Shackelford (Eds.), *The Oxford handbook of comparative evolutionary psychology* (pp. 274–291). Oxford, UK: Oxford University Press. doi:10.1093/oxfordhb/9780199738182.013.0015
- Smith, J. D., Coutinho, M. V. C., Church, B. A., & Beran, M. J. (2013). Executive-attentional uncertainty responses by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *142*, 458–475. doi:10.1037/a0029601
- Smith, J. D., & Minda, J. P. (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. doi:10.1037/0278-7393.27.4.984
- Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Monkeys adaptively monitor uncertainty while multi-tasking. *Animal Cognition*, *13*, 93–101. doi:10.1007/s10071-009-0249-5
- Smith, J. D., Redford, J. S., & Haas, S. M. (2008). Prototype abstraction by monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: General*, *137*, 390–401. doi:10.1037/0096-3445.137.2.390
- Smith, J. D., Shields, W. E., Allendoerfer, K. R., & Washburn, W. A. (1998). Memory monitoring by animals and humans. *Journal of Experimental Psychology: General*, *127*, 227–250. doi:10.1037/0096-3445.127.3.227
- Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. *Cognition*, *62*, 75–97.
- Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *The Behavioral and Brain Sciences*, *26*, 317–339. doi:10.1017/S0140525X03000086
- Staddon, J. E. R., Jozefowicz, J., & Cerutti, D. (April 13, 2007). Metacognition: A problem not a process. *PsyCrit*, 1–5.
- Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition*, *11*, 21–42.
- Sutton, J. E., & Shettleworth, S. J. (2008). Memory without awareness: Pigeons do not show metamemory in delayed matching-to-sample. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*, 266–282. doi:10.1037/0097-7403.34.2.266
- Teller, S. A. (1989). *Metamemory in the pigeon: Prediction of performance on a delayed matching to sample task* (Undergraduate thesis). Reed College, Portland, Oregon.
- Templer, V. L., & Hampton, R. R. (2012). Rhesus monkeys (*Macaca mulatta*) show robust evidence for memory awareness across multiple generalization tests. *Animal Cognition*, *15*, 409–419. doi:10.1007/s10071-011-0468-4
- Washburn, D. A., Gullledge, J. P., Beran, M. J., & Smith, J. D. (2010). With his memory magnetically erased, a monkey knows he is uncertain. *Biology Letters*, *6*, 160–162. doi:10.1098/rsbl.2009.0737
- Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 185–189. doi:10.1037/0097-7403.32.2.185
- Zentall, T. R., & Stagner, J. P. (2010). Pigeons prefer conditional stimuli over their absence: A comment on Roberts et al. (2009). *Journal of Experimental Psychology: Animal Behavior Processes*, *36*, 506–509. doi:10.1037/a0020202

(Appendix follows)

Appendix

An Associative Model Illustrated

Le Pelley’s (2012) model considered how animals might learn associatively to make adaptive URs in uncertainty tasks. His model grants the organism an array of response-strength nodes as follows.

The array is indexed by a series of perceptual impressions (e.g., the increasing level of perceived trial density across trial levels in levels in Smith et al., 2006). Le Pelley’s (2012) model borrows from signal-detection theory the correct idea that animals (and humans) respond to their subjective impressions of trial levels—not to the objective trial levels. This is the model’s way to incorporate the inevitable possibility of perceptual error. The assumption is that the objective trial level may be misperceived. So, a Level 20 could be misperceived as Level 24. Or, a Level 32 could be misperceived as 28. The subjective impressions are the

result of objective stimuli filtered through perceptual error. In Le Pelley’s model and many others, the observer is assumed to act on the basis of its subjective impressions because it cannot experience directly the objective stimulus level in itself.

The array is also indexed by the simulated performer’s three possible responses (e.g., Sparse, Dense, and Uncertain in Smith et al., 2006). The array contains response-strength activations: The simulated performer with a subjective impression of 12 might find the highest response strength at Sparse and respond Sparse; the simulated performer with a subjective impression of 34 might find the highest response strength at Dense and respond Dense. The array is updated based on trial outcomes at the array position of the subjective impression felt on that trial and the response made on

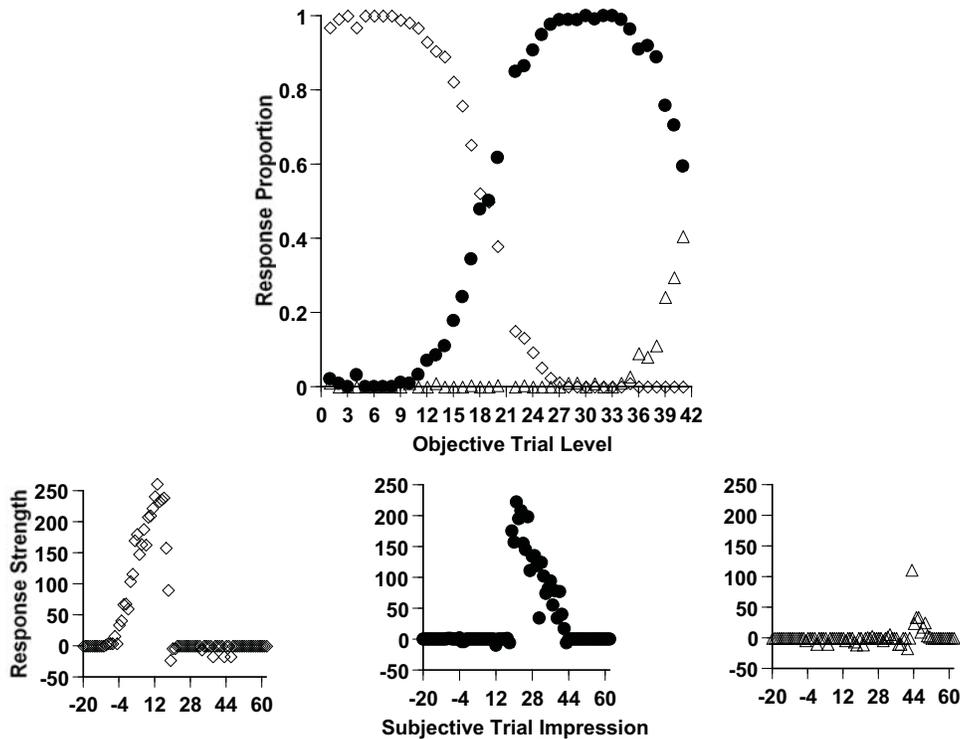


Figure A1. Top. The performance of a simulated subject in the paradigm of Smith et al. (2006), as modeled by Le Pelley’s (2012) associative-learning model. The horizontal axis indicates the density level of the box. The Sparse and Dense responses, respectively, were correct for boxes at density Levels 1–20 and 22–41. The open diamonds, filled circles, and open triangles show, respectively, the proportion of Sparse, Uncertain, and Dense responses. Bottom. An X-ray of the contents of the simulated subject’s associative array, containing response strengths for Sparse (left), Uncertain (middle), and Dense (right) responses at subjective trial impressions of trials that run from –20 (trials making an intensely sparse impression) to 60 (trials making an intensely dense impression).

(Appendix continues)

that trial that caused the outcome—that is, at the array position (Subjective Impression, Response Made). For positive outcomes, the response strength at that subjective impression would be incremented and become more likely to control responding on similar trials in the future. For negative outcomes, the response strength at that subjective impression would be decremented and become less likely to control responding on similar trials in the future. In this way, the array's activations come to reflect the reinforcement history in the task. The array tends to align with the task's associative structure, and this alignment can help the simulated performer behave more adaptively.

Figure A1 (top panel) illustrates one possible predicted data pattern by Le Pelley's (2012) model over 6,000 trials in a Sparse–Dense discrimination. This simulation was conducted using the same parameter settings as Le Pelley. Le Pelley has correctly pointed out that his model is highly unstable and variable, and this remains a significant problem in its interpretation. In this particular case, the particular learning history produced predominant Sparse responding for trial Levels 1 to 18, then a dominant UR tendency spanning the rest of the stimulus continuum (trial Levels 18–41). Additional simulation work in our laboratory has shown that Le Pelley's model is subject to overwhelming effects caused by just a few accidents at the very beginning of training of which trial kinds and levels are presented. Le Pelley discussed at length why this remains a problem with his associative model and the steps that might be taken to remedy it.

An interesting feature of Le Pelley's (2012) model is that it lets one examine the associative register's contents, indexed as just described by the simulated observer's subjective impression of trials. The response-strength registers are shown from left to right, for Sparse, Uncertain, and Dense responses, respectively, in the bottom panels of Figure A1. Subjective impressions of trials can occur that are below Level 1, or above Level 41, as trials feel especially Sparse (because perceptual error acts downward) or Dense (as perceptual error acts upward) to the simulated performer. At left, the tendency to respond Sparse is qualitatively dominant up to trial Level 18, even though those trials are very difficult trials that real animals might feel conflicted about. For these trial levels, the response strength for the UR is constant and effectively 0, though one might suppose that UR response tendencies would grow as trials grow more difficult. Then there is a qualitatively sudden transition in which the UR response tendency becomes overwhelmingly dominant. Appropriately, this response strength wanes until, for extreme subjective impressions, the Dense response begin to gain some activation traction (right panel).

Received September 16, 2012

Revision received April 12, 2013

Accepted April 15, 2013 ■

E-Mail Notification of Your Latest Issue Online!

Would you like to know when the next issue of your favorite APA journal will be available online? This service is now available to you. Sign up at <http://notify.apa.org/> and you will be notified by e-mail when issues of interest to you become available!