

I Scan, Therefore I Decline: The Time Course of Difficulty Monitoring in Humans (*Homo sapiens*) and Macaques (*Macaca mulatta*)

J. David Smith
Georgia State University

Joseph Boomer
University at Buffalo, The State University of New York

Barbara A. Church
Georgia State University

Alexandria C. Zakrzewski
University of Richmond

Michael J. Beran
Georgia State University

Michael L. Baum
University at Buffalo, The State University of New York

The study of nonhumans' metacognitive judgments about trial difficulty has grown into an important comparative literature. However, the potential for associative-learning confounds in this area has left room for behaviorist interpretations that are strongly asserted and hotly debated. This article considers how researchers may be able to observe animals' strategic cognitive processes more clearly by creating temporally extended problems within which associative cues are not always immediately available. We asked humans and rhesus macaques to commit to completing spatially extended mazes or to decline completing them through a trial-decline response. The mazes could sometimes be completed successfully, but other times had a constriction that blocked completion. A deliberate, systematic scanning process could preevaluate a maze and determine the appropriate response. Latency analyses charted the time course of the evaluative process. Both humans and macaques appeared, from the pattern of their latencies, to scan the mazes through before committing to completing them. Thus monkeys, too, can base trial-decline responses on temporally extended evaluation processes, confirming that those responses have strategic cognitive-processing bases in addition to behavioral-reactive bases. The results also show the value of temporally and spatially extended problems to let researchers study the trajectory of animals' online cognitive processes.

Keywords: associative learning, cognitive processing, comparative cognition, primates, visual search

Humans' metacognition is fundamental to their thinking and reflective decision-making. An extensive literature explores metacognition (Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2008; Koriat & Goldsmith, 1994; Nelson, 1992) and its development (Balcomb & Gerken, 2008; Brown, Bransford, Ferrara, & Campi-

one, 1983; Flavell, 1979). Metacognition is a sophisticated cognitive capacity possibly linked to consciousness (Koriat, 2007; Nelson, 1996). Metacognitive assessments often occur deliberately and reflectively, probably using working memory processes (Paul et al., 2015; Schwartz, 2008). Some have judged metacognition so sophisticated as to be uniquely human (Metcalf & Kober, 2005).

Yet some nonhumans might share aspects of a metacognitive capacity. If so, then metacognitive paradigms could show the roots of reflective cognition in nonhuman animals (hereafter, animals), informing theoretical debates in comparative psychology. Accordingly, researchers have explored animal metacognition extensively (reviews in Kornell, 2009; Metcalfe, 2008; Smith, Beran, & Couchman, 2012; primary research in Basile, Schroeder, Brown, Templer, & Hampton, 2015; Beran & Smith, 2011; Call, 2010; Couchman, Coutinho, Beran, & Smith, 2010; Foote & Crystal, 2007; Fujita, 2009; Kornell, Son, & Terrace, 2007; Paukner, Anderson, & Fujita, 2006; Roberts et al., 2009; Smith, Coutinho, Church, & Beran, 2013; Suda-King, 2008; Sutton & Shettleworth, 2008; Templer & Hampton, 2012; Washburn, Gullede, Beran, & Smith, 2010; Zakrzewski, Perdue, Beran, Church, & Smith, 2014—these references sample a large literature produced by many researchers). Nonhuman primates have produced apparently

This article was published Online First April 16, 2018.

J. David Smith, Department of Psychology and Language Research Center, Georgia State University; Joseph Boomer, Department of Psychology, University at Buffalo, The State University of New York; Barbara A. Church, Language Research Center, Georgia State University; Alexandria C. Zakrzewski, Department of Psychology, University of Richmond; Michael J. Beran, Department of Psychology and Language Research Center, Georgia State University; Michael L. Baum, Department of Psychology, University at Buffalo, The State University of New York.

The preparation of this article was supported by Grants HD061455 and HD060563 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and Grant BCS-0956993 from National Science Foundation.

Correspondence concerning this article should be addressed to J. David Smith, Department of Psychology, Georgia State University, 738 Urban Life Building, Atlanta, GA 30303. E-mail: jsmith395@gsu.edu

metacognitive performances. That is, they have often shown that they can selectively and adaptively make trial-decline responses facing difficult trials that might cause error, using those responses to erase the difficult trial and move on to the next, randomly scheduled trial. Perhaps some animals do reflectively evaluate their task prospects and regulate their behavioral choices accordingly. These prospective evaluative processes are the focus of the present article. Given our history as animal-metacognition researchers, it is useful for us to say directly that these evaluative processes may or may not be metacognitive, and in the context of the present article, this does not matter.

The Associative Perspective

However, the standard is strict for concluding that these evaluative processes exist, because comparative psychology traditionally has supported low-level interpretations of animal behavior when these are viable (Morgan, 1906). Instead of concluding for deliberate cognitive processes, one could emphasize associative processes. In fact, animals' trial-decline responses might be explained as conditioned reactions if they are cued by associative stimuli or entrained by reinforcement. Thus, this literature has fostered strong and debated assertions of associative-learning processes (Basile & Hampton, 2014; Basile et al., 2015; Carruthers, 2008; Carruthers & Ritchie, 2012; Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009a, 2009b; Le Pelley, 2012, 2014; Smith, 2009; Smith et al., 2012; Smith, Beran, Couchman, & Coutinho, 2008; Smith, Couchman, & Beran, 2012, 2014a, 2014b; Staddon, Jozefowicz, & Cerutti, 2007).

Moreover, these assertions have force, because many paradigms rely on difficult, ambiguous stimuli to engender uncertainty and to foster trial-decline responses. But these difficult stimuli are also associated during training with errors and scant rewards. Associative-learning processes could have been shaped by this contingency, conditioning low-level trial-avoidance responses to problematic stimuli. These responses would seem metacognitive, but they might not be. Even the inaugural studies of animal metacognition included this potential associative confound (Smith et al., 1995; Smith, Shields, Schull, & Washburn, 1997).

One sees that attributing sophisticated difficulty monitoring to animals is a difficult matter. In fact, Metcalfe (2008) argued that no task that makes stimuli immediately available to animals can be judged to be a test of their metacognition. Available stimuli can always function as associative cues to which animals show low-level, conditioned reactions. Hampton's (2009) insightful analysis of animal-metacognition paradigms also focused on the availability of associative cues and on their potential confounding effects in experimental designs.

Researchers have responded to potential confounds by seeking to eliminate them. For example, Shields, Smith, and Washburn (1997) sought to elevate animals' difficulty monitoring above the plane of associative cues and reactible stimuli. Shields et al. gave macaques difficult trials in a same-different relational-judgment task. This task requires an abstraction beyond the absolute stimuli that carry the relation. In this sense, the absolute stimuli do not matter. The abstractness explains why true same-different performances are phylogenetically restricted and cause even nonhuman primates great difficulty (Flemming, Beran, & Washburn, 2007; Katz, Wright, & Bachevalier, 2002; Premack, 1978; Smith, Flem-

ming, Boomer, Beran, & Church, 2013). The logic of Shields's paradigm was that if animals do decline difficult relational judgments, then perhaps they are showing higher level assessments of difficulty. However, some conceptions of associative theory might accept that an abstract relation is still a stimulus in a sense that might let it serve as an associative cue (Debert, Matos, & McIlvane, 2007).

For another example, Hampton (2001) sought to move animals' difficulty monitoring inward, away from concretely available stimuli, so that trial-decline responses would reflect memory monitoring, not stimulus reaction. In a delayed-matching-to-sample task, rhesus macaques decided whether to accept or decline tests of recently presented memory material. This decision best depended on whether they had an active trace of the to-be-matched sample available in working memory. The logic of this paradigm was that if animals declined trials facing faint memories, then perhaps they were showing higher level memory monitoring. However, some conceptions of associative-learning theory might still embrace an internal memory trace as a stimulus that can trigger behavior in low-level, associative ways (Le Pelley, 2012). It is difficult to eliminate associative cues entirely from the tasks given to animals.

A New Approach

In this article, we take a complementary approach to this problem. We arranged our task so that the cues that would be useful to difficulty monitoring were not immediately available to the animal. We reasoned that if these cues were not immediately available, they could not be reacted to associatively until they were perceived and processed. We arranged matters so that the relevant cues to trial difficulty would only become available at the end of a deliberate, systematic search by the animal. We reasoned that if the situation required a systematic strategy, then one could know that trial-decline responses were the outcome of that kind of strategy, and not solely the outcome of a low-level, associative reaction. The essence of our paradigm was to create a situation in which humans' and animals' difficulty monitoring was stretched so that it necessarily encompassed the evaluation of temporally extended problems. Here, too, we stress that this article is not about proving metacognition, or about insisting that our paradigm is a new instance of animal metacognition. Instead, we are exploring an empirical means by which one may constructively move beyond purely associative interpretations of animals' performances, toward studying and interpreting their higher level cognitive processes.

We can give an analogy of what we will try to do here. There is a dreaded response alternative in national achievement tests that says the following: e) the answer cannot be determined from the information given. This response isn't given associatively or reactively (at least not adaptively so!). It is given as the result of cognitive processing invested until that processing reaches a judgment of indeterminacy. Similarly, though within monkeys' behavioral/cognitive limits, we sought to make the trial-decline response a behavioral declaration—following a deliberate, systematic search strategy—that the trial could not be completed based on the information given.

A less lofty analogy highlights other aspects of our task. Suppose a parent constructs a code for a child to break revealing where a Hershey bar is hidden. Now one can observe the child's code-

breaking processes in the absence of available associative stimuli. Of course, in the end, there is an associative stimulus—the Hershey bar!—that will be reacted to associatively (it is a Hershey bar!). But that doesn't matter, in this example or in our task. What matters is that the temporal extension and separation in the situation opens a window on cognitive processing that is not associatively governed. Our goal was to open this window on animals' systematic search strategies, by distancing and separating available associative stimuli. An important subtext of our article is that one needn't view tasks dichotomously as either being associative or higher level. Many tasks probably have both elements. We are trying to show that, given a mixed situation, there are empirical methods that may constructively separate different processes so that scientists and theorists may see them more clearly.

To be precise, we presented to humans and animals both Possible and Impossible maze-like stimuli composed of a spatially extended series of wicket gates. Half of the trials were Possible to complete. On these trials, all of the wickets were generously wide enough to let the participant's self-controlled cursor pass through, continuing on to finally reach the maze's goal position. For these mazes, it was correct for participants to commit to completing the maze. But half the maze displays were Impossible. On these trials, just one of the many wickets presented a constriction that would not allow the cursor to continue. For these mazes, the better behavioral choice was to decline the trial and move on to the next randomly scheduled trial.

One possible outcome of this experiment is that the narrow wickets would “pop out” from the visual display. Then one would see accurate performance and short response latencies no matter the position of the wicket in the maze. Another possible outcome, especially for the monkeys, is that subjects might take one fixation at one part of the array and respond based on that. Then one would see poor performance and short latencies no matter the narrow wicket's spatial position. The third possibility is the most interesting. The narrow wickets might not pop out, and subjects might not react associatively to one glance. Instead they might search the displays completely for the narrow wicket that might or might not be there. Most efficiently of all, they might search systematically—for example, in left-to-right spatial order—so that their response latencies would reflect spatial position lawfully. To evaluate this possibility, we used reaction time (RT) analyses to help us understand the character of humans' and monkeys' scanning and evaluation processes.

Experiment 1A: Humans

Method

Participants. Thirty-four undergraduates with normal or corrected vision participated in Experiments 1A to fulfill a course requirement. We applied these performance criteria: (a) Participants included for analysis had to have completed at least 80 trials, providing sufficient trials for analysis; (b) they had to have declined at least 10% of trials, allowing us to assess the distribution of the trial-decline response across Possible and Impossible trials and across different positions of the constriction point within the maze; (c) they had to have completed (not declining) at least 10% of trials, allowing us to assess the distribution of their decision to accept trials across those trial types. The data from eight of 34

participants were excluded on these bases—five owing to underusing the trial-decline response, two owing to completing too few trials, and one owing to overusing the trial-decline response. In Experiment 1A, the data from 26 participants were included for analysis.

Maze stimuli. The crucial stimulus (see Figure 1) was a series of wickets stretching across the computer screen (one wicket, or gate, every 20 pixel positions). Nearly all of the wickets were at least 15 pixels wide, allowing generous room for our 10-pixel cursor to fit between them and to progress through the maze. Precisely, the wickets varied randomly from 15 to 20 pixels wide, providing the maze display more visual complexity. Moreover, the wickets were randomly placed up and down on the screen, across 30 random pixel positions, giving the display the look of a slalom course, adding additional visual complexity, and ruling out the apprehension of the gate structure in one glance.

On Impossible trials, one of the gates in the range of Gate 4–Gate 23 was 10 pixels wide, creating a constriction point that the cursor could not pass. We did not put constriction points near the display's start or finish where they might be immediately perceived and defeat the deliberate scanning process we hoped to foster. The stimuli were shown on a 17-in. monitor with 800×600 pixel resolution and viewed from about 24 in. The maze itself, stretching about 12 in. across the screen, was quite spatially extended, subtending a visual angle of about 29 degrees. This spatial extent also fostered systematic scanning patterns because the maze displays could not be entirely apprehended all at once.

Maze trials. Each trial was dominated by the maze stimulus, which was 26 gates separated by 20 pixel positions and spanning the computer screen. Maze stimuli were Possible or Impossible, with half the trials of each type. The trial type on each trial was decided truly randomly.

To the right of the 26th wicket of the maze, a blue circle spanning 10 pixel positions marked the goal state of maze completion. To the left of the first wicket of the maze, a red circle spanning 10 pixel positions was the response cursor that could be



Figure 1. Illustrating an Impossible maze trial. Counting from the left, the 21st gate or wicket has narrowed, creating a constriction point that cannot be negotiated by the red cursor that the subject controlled. Impossible maze trials required a trial-decline response (?) to fend off a long penalty time-out. In contrast, on Possible maze trials, all 26 wickets were generously wide for passage by the cursor to the blue-circle goal position. See the online article for the color version of this figure.

controlled by the participant. Far left on the screen, there was a large-font?, indicating the position to move the cursor to make a trial-decline response.

Humans used the S and L keyboard keys to produce movements of the red cursor toward the? (leftward, S keyboard presses) or toward the maze and then through it (rightward, L key presses). It required about 15 key presses or cursor movements to reach the?, completing a trial-decline response. It required about 15 key presses or cursor movements to enter the maze, thus committing to trying to complete it. The S and L keys could be held down so that these multiple key presses registered automatically. There was substantial room and time as the cursor moved right or left for the participant to think, reflect, and *rethink* during the movement toward the ? or the maze entrance. This is important because the task had a catch. If one approached too closely the first gate of the maze (i.e., within 10 screen pixels of it), one committed to attempting the maze, whether this approach had been intentional or not. From that point the red cursor would only move to the right, to the goal position on Possible trials but only up to the constriction point on Impossible trials. The idea behind this point-of-no-return was to place a premium on foresight and the preevaluation of the maze. Once the commitment to complete the maze was made, the actual completion of the maze was trivial. The cursor self-centered through the gates as it progressed from Left to Right—all the participant needed to do was to depress the L key to sustain the rightward progression.

Participants reaching the goal state on a Possible trial received 3 points added to their total, as indicated to them on the screen in green text. Participants blocked at a constriction point lost 3 points, as indicated to them in red text. Each point loss was accompanied by a 7-s penalty time-out, so 21 s were lost for each Impossible trial committed to wrongly.

Participants making the trial-decline response did not receive or lose points or suffer any time consequence. The present maze stimulus simply vanished, replaced with the next random maze trial chosen by the computer. This new trial could once again be Possible or Impossible, in the latter case with a constriction point in a new randomly chosen position. The maze task continued for 55 min or 600 trials, whichever came first.

Instructions. Participants were told that their goal was to move the red cursor across the screen to touch the blue circle. They were told that on some trials, the path might grow too narrow to reach the goal successfully, and that on those trials they should move the cursor to the? to skip the trial and avoid an error. They were told that on other trials, the path would stay wide enough to let them reach the goal, so that they could complete the trial to win points. It was emphasized to them that they should complete the trials that they could successfully, but respond? for the paths that would fail, so as to maximize points, minimize errors, and avoid time-outs.

Results

Humans completed 4,177 trials (about 160 per participant), 2,074 Possible trials and 2,103 Impossible trials with one of the 20 gates (4–23) constricted. On Possible trials, they declined 277 trials, or 13%. On Impossible trials, they declined 1,640 trials, or 78%. So, they were six times more likely to decline Impossible trials than Possible trials. They committed to just 463 Impossible

trials. Across both trial types, they made the appropriate response on 82% of trials.

The crucial result concerns participants' trial-decline latencies. If these latencies reflected a systematic evaluative prescan of the maze, a scan with directionality, then we should see orderly changes in average latency across the gates of the maze. To evaluate this possibility, we analyzed Impossible trials only, and only trials that finally earned the decline response. These are the trials/responses that may have reflected an effective scanning process, with the constriction point registered and avoided. These choices would then be slower as the constricted gate was farther right—if the participant scanned from left to right. These choices also might be slower as the constricted gate was farther left—if the participant scanned from right to left. Preliminary to this analysis, we clipped the RT distribution by excluding any latency greater than 15 s. These latencies were deemed not to reasonably reflect a veridical scanning latency, but rather an off-task moment, and so those trials were not included. With these exclusions made—34 latencies, or 2.1% of the 1,640 relevant trials—we analyzed 1,606 latency events, about 80 events for each of the 20 possibly constricted gates in the maze.

Figure 2A shows the mean latency to a trial-decline response plotted against Maze Gates 4–23, counting Maze Gates from left to right across the screen. Trial-decline latencies increased from 3.9 s to 6.1 s across the gates. (Because we trimmed the RT distributions to exclude skewing outliers, this data pattern remained similar if we used median instead of mean latencies. Across the 20 gate positions, mean and median latencies only differed by 0.24 s on average.) To test the reliability of the pattern of mean latencies, we compared humans' latency for Gates 4–8 (386 observed trial latencies for these gates) with that for Gates 19–23 (380 observed trial latencies for these gates). These latencies were 3.87 s, 95% confidence interval (CI) [3.64, 4.10], and 5.56 s, 95% CI [5.23, 5.89], respectively, standard deviations of 2.28 and 3.27, $t(764) = 8.26$, $p < .001$, $d = .597$.¹

Discussion

All the RT patterns suggest that humans systematically scanned the spatially extended arrays to decide whether mazes were Possible or Impossible. Their decision latencies grew systematically longer as the constriction point lay farther down the maze. The result is conservative, because we gave up 25% of the maze gates at the start and the finish to make our constriction points less salient, and because the analysis would be weakened if all participants did not adopt the same left-to-right scanning pattern.

One can see why the participants did adopt a systematic scanning strategy. Many aspects of the task (e.g., the up/down scatter of the gates, the width variation in gates, the absence of constriction points early and late in the maze, and the number of gates in the maze) ruled against participants' seeing constrictions as pop-out events and against their apprehending the maze in one glance.

¹ Trial analyses are reported here to maximize comparability with the later monkey results. However, a standard subject analysis comparing latencies when constrictions appeared in the first half versus the last half of the wickets was also significant, $t(25) = 5.933$, $p < .001$, $d = .731$, as was a general linear model (GLM) using constriction position as the independent variable, $F(19, 190) = 4.336$, $p < .001$, $\eta_p^2 = .302$.

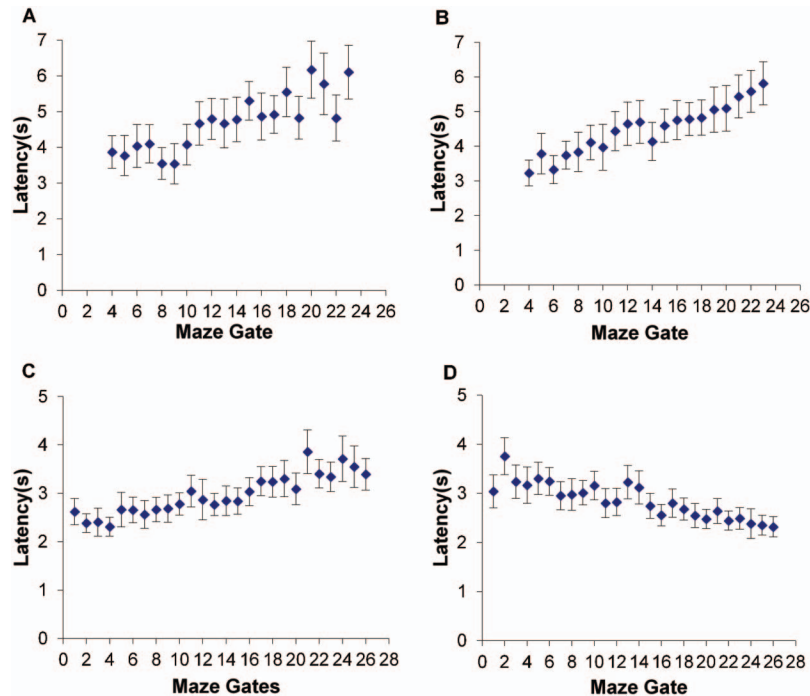


Figure 2. (A) Humans' mean latency for trial-decline responses on Impossible maze displays in Experiment 1A. These responses depended on the detection of a single constriction somewhere along the run of the maze. The participant-controlled cursor could not pass this constriction. Each data point is framed by its 95% confidence interval. (B–D) Humans' mean latency for trial-decline responses on Impossible trials in Experiments 1B–D, depicted in the same way. Panels C and D reflect performance with maze displays containing highly visible wicket gates. See the online article for the color version of this figure.

Participants did strongly need to scan across the maze, and the RTs indicate that they did so.

We point out that our maze task probably depends on both higher level cognitive processes and low-level associative processes. This acknowledgment is especially appropriate in a cross-species article like this, but it could be appropriate even in discussing purely human research. On the one hand, participants' systematic and deliberate scans across the maze would represent a sophisticated and high-level cognitive process. Their scans are not triggered by, or reactive to, either positive or negative stimuli in the spatially extended arrays, because any positive or negative stimuli in the arrays have not been "seen" or perceptually registered yet (or else the systematic scan would not be necessary). These stimuli do not pop out, and they will not be registered until the participant comes to them during the preevaluation of the maze. If they did pop out or impress themselves automatically and immediately in some other way, then participants' maze preevaluations would not show the clear relationship between response latency and wicket position. In this article, when we refer to the sophisticated scanning strategies of subjects preevaluating maze displays, we are referring to this higher level cognitive process.

On the other hand, closed maze gates may well come to be slightly aversive to participants and avoided, possibly reactively and according to traditional principles of associative learning. So clearly there are also associative cues and associative reactions in our task. Thus, we are not precluding that our task contains associative elements. We are not insisting that our task disproves

associative learning or provides evidence against it. Instead, we suggest that our task contains cognitive elements of visual search that complement or supplement the associative elements. Most important, the temporal and spatial extension of our maze trials accomplished the constructive goal of separating these stages of information processing so that they became observable independently.

Perhaps all tasks have elements of both kinds of processing within them. We are comfortable with that possibility. In all of our work, we are trying to find the appropriate theoretical balance between these two important forms of information processing. Here, we tried to study that balance using the process segregation allowed by temporally extended displays.

Experiments 1B–D: Experimental Variations With Humans

We conducted more studies with humans. These were designed to inform our approach to testing rhesus macaques in equivalent procedures.

Experiment 1B

Participants. We tested 45 humans drawn from the same participant population and subjected to the same exclusion criteria discussed in Experiment 1A. Eight were excluded for not completing enough trials. Thirty-seven participants were included in the analyses.

Procedure. Many aspects of the task in Experiment 1B were identical to those already described in Experiment 1A. We made several changes to sharpen our data pattern by motivating participants to “do the right thing”—that is, to scan systematically across the maze stimuli. First, we increased the time-out to 30 s for committing to an Impossible trial, a serious but completely avoidable time-out. Second, we gave participants forced experience with all the task’s responses. On 10% of trials, the trial-decline response was forced because the cursor would not move rightward, letting them feel this contingency. On 10% of trials, the commitment to complete the maze was forced because the cursor would not move leftward, letting them feel this contingency. Participants chose either response optionally on the remaining 80% of trials. Third, we instituted a mandatory reflection time of 1 s as each maze trial appeared, thus denying participants impulsive responses. We note that this approach has a potential downside. Participants might accomplish part of their scan gratis during the 1 s that would not be included in their overall scanning time.

Results. The 37 participants completed 3,564 optional trials in the task (both kinds of forced trials were uninteresting to analyze), 1,784 Possible trials and 1,780 Impossible trials with one of the 20 gates (4–23) constricted. On Possible trials, they declined 55 trials, or 3%. On Impossible trials, they declined 1,482 trials, or 83%. They were 27 times more likely to decline Impossible trials than Possible trials. They committed to just 298 Impossible trials. Across both trial types, they made the appropriate response on 90% of trials.

Figure 2B shows the mean latency to a trial-decline response plotted across maze gates left to right across the screen. Trial-decline latencies increased from 3.2 s to 5.8 s across the gates. To test the reliability of this pattern, we compared humans’ latency for Gates 4–8 (377 observed trial latencies for these gates) with that for Gates 19–23 (361 observed trial latencies for these gates). These latencies were 3.58 s, 95% CI [3.37, 3.79], and 5.41 s, 95% CI [5.13, 5.69], respectively, standard deviations of 2.07 and 2.69, $t(736) = 10.30, p < .001, d = .761$.² The measured latencies were faster now than in Experiment 1A, perhaps because some scanning was accomplished during the 1-s delay imposed to prevent impulsive responding.

Experiment 1B did sharpen our data pattern, by motivating participants more strongly, by granting them experience in all task contingencies, and perhaps by fostering reflection. It also perfectly replicated Experiment 1A. None of the methodological adjustments in Experiment 1B affected the character of the outcome.

Experiment 1C

Participants. We tested 19 humans drawn from the same participant population and subjected to the same exclusion criteria discussed in Experiment 1A. None met the criteria for exclusion.

Procedure. Many aspects of the task in Experiment 1C were identical to those already described in Experiment 1A. In Experiment 1C, we made two adjustments. First, we experimented with gates composed of larger and more visible endpoints. Now the wicket gates were defined by circles of radius 4 pixel positions. We understood that we might need to make the slalom gates more visible to monkeys. Second, we used all 26 wickets as possible constriction points. We understood that we might need to use

smaller maze arrays with monkeys, and that therefore we might need to include all gates.

Results. The 19 participants completed 5,514 trials, 2,769 and 2,745 Possible and Impossible trials, respectively. On Possible trials, they declined 166 trials, or 6%. On Impossible trials, they declined 2,274 trials, or 83%. So, they were 14 times more likely to decline Impossible trials than Possible trials. They committed to 471 Impossible trials. Across both trial types, they made the appropriate response on 88% of trials.

Figure 2C shows the mean latency to a trial-decline response plotted against Maze Gates 1–26. Trial-decline latencies increased from 2.6 s to 3.4 s across the gates. To test the reliability of this pattern, we compared humans’ latency for Gates 1–5 (451 observed trial latencies for these gates) with that for Gates 22–26 (444 observed trial latencies for these gates). These latencies were 2.47 s, 95% CI [2.35, 2.59], and 3.48 s, 95% CI [3.32, 3.64], respectively, standard deviations of 1.273 and 1.758, $t(893) = 9.84, p < .001, d = .658$.³

Experiment 1C was still a strong and robust replication of the basic scanning-strategy phenomenon. The larger, salient wickets did speed up scanning, without changing the result. Including all maze gates as possible constriction points—the whole run of the maze—did not prove problematic.

Experiment 1D

Participants. We tested 31 humans drawn from the same participant population and subjected to the same exclusion criteria discussed in Experiment 1A. Three were excluded for failing to use the trial-decline response at least 10% of the time. Twenty-eight participants were included in the analyses.

Procedure. Many aspects of the task in Experiment 1D were identical to those already described in Experiment 1A. But Experiment 1D did differ in one dramatic aspect. The left-to-right scanning strategy is not the only efficient strategy in our task, though it was salient to us in our human-researcher planning and apparently in human-participant performance, too. We wondered how ably humans would scan from right to left instead. We realized that monkeys might well do so, having no reading bias. So, we examined human’s right-to-left scanning proficiency. In Experiment 1D, the? was at the far right of the screen, with the red cursor and the beginning of the maze. So, the performance of the maze would now occur in right-to-left fashion, and participants might therefore scan right to left, though they could still obey their left-to-right bias by scanning from the blue goal circle rightward.

Results. The 28 participants completed 8,925 trials, 4,462 and 4,463 Possible and Impossible trials, respectively. Participants made the trial-decline response on 14% of Possible trials. They made the trial-decline response on 92% of Impossible trials—six

² A standard subject analysis comparing latencies when constrictions appeared in the first half versus the last half of the wickets was also significant, $t(36) = 7.125, p < .001, d = .915$. A standard GLM using constriction position as the independent variable was not significant, $F < 2$.

³ A standard subject analysis comparing latencies when constrictions appeared in the first half versus the last half of the wickets was also significant, $t(18) = 5.923, p < .001, d = .952$, as was a GLM using constriction position as the independent variable, $F(25, 275) = 6.360, p < .001, \eta_p^2 = .366$.

times as often. Overall, they made the appropriate response on 89% of trials.

Figure 2D shows the mean latency to a trial-decline response plotted against Maze Gates 1–26. Trial-decline latencies increased from 2.3 s at Gate 26 (farthest right) to 3.0 s at Gate 1 (farthest left). We tested the reliability of this pattern as before. The early latencies (809 observed trial latencies) and late latencies (799 observed trial latencies) were 2.39 s, 95% CI [2.29, 2.50], and 3.31 s, 95% CI [3.15, 3.46], respectively, standard deviations of 1.455 and 2.264, $t(1606) = 9.58$, $p < .001$, $d = .479$.⁴ This pattern of RTs suggests that humans did reverse the direction of their scanning to comport with the mirror reflection of the task, and they revealed no disability in making this reversal. Throughout Experiments 1A–D, humans showed the same strong suggestion of this systematic spatial scanning strategy, using a directional preevaluation of the maze displays as Possible or Impossible. Indeed, it is difficult to account for the RT patterns in any other way. Experiment 2 explores the performance of two rhesus monkeys in highly similar paradigms.

Experiment 2: Rhesus Macaques (*Macaca mulatta*)

Method

Participants. Male macaques (*Macaca mulatta*) Lou and Obi (21 and 11 years old, respectively) were tested. They had been trained as described elsewhere (Washburn & Rumbaugh, 1992) to respond to computer-graphic stimuli by manipulating a joystick. They had participated in previous computerized experiments (Beran, Evans, Klein, & Einstein, 2012; Beran, Perdue, & Smith, 2014; Beran & Smith, 2011; Smith, Redford, Beran, & Washburn, 2010). Thus, as in many other comparative studies, our subjects were cognitively experienced, not cognitively naïve. In fact, here it was crucial to test cognitively experienced animals because they would show the top of monkeys' capacity in this area and provide the best indication of possible cognitive continuities with humans. Macaques were tested in their home cages at the Language Research Center (Georgia State University), with ad lib access to the test apparatus, working when they chose during long sessions. They had continuous access to water. They worked for fruit-flavored primate pellets. They received a daily diet of fruits and vegetables independent of task participation, and thus, they were not food deprived for the purposes of this experiment.

This study exemplifies a venerable tradition within comparative psychology of conducting intensive empirical investigations with smaller numbers of animal participants. Here, the two monkeys completed 145,583 trials to produce the experiment's data. Small-sample research has played a crucial role in comparative psychology's empirical success and theoretical development. For example, it has anchored the fields of ape language, parrot cognition, dolphin language, apes' conceptual functioning, apes' theory of mind, self-awareness, metacognition, and other fields as well (e.g., respectively, Savage-Rumbaugh, 1986; Pepperberg, 1983; Herman & Forestell, 1985; Boysen & Berntson, 1989; Premack & Woodruff, 1978; Gallup, 1982; Smith et al., 1995). This approach suited well our present research purposes.

Apparatus. The monkeys were tested using the Language Research Center's Computerized Test System (Washburn & Rumbaugh, 1992), comprising a computer, a digital joystick, a color

monitor, a pellet dispenser, and programming code written in Turbo Pascal 7.0. Trials were presented on a 17-in. color monitor with 800 × 600 resolution. Monkeys viewed the stimuli from a distance of about 2 ft (60.96 cm). Joystick responses were made with a Logitech Precision gamepad, which was mounted vertically to the test station. Monkeys manipulated the joystick, which protrudes horizontally through the mesh of their home cages, producing isomorphic movements of a computer-graphic cursor on the screen. Contacting the goal with the cursor brought them a 94-mg fruit-flavored chow pellets (Bio-Serve, Frenchtown, NJ) using a Gerbrands 5120 dispenser interfaced to the computer through a relay box and output board (PIO-12 and ERA-01; Keithley Instruments, Cleveland, OH).

Trials. The monkeys' tasks were similar to those already described, with just some species-specific modifications described now. We used the left-to-right version of the task as used in Experiments 1A–C. We used the large and salient gate format as described in Experiments 1C and D. Except where indicated later in the text, we instituted the brief pause-cursor procedure also used with humans to prevent unintentional inertial or impulsive Left or Right cursor movements. The monkeys are familiar with this procedure, as it is common in our research with them. We varied the number of gates in the maze displays, in particular increasing the spatial extent of the maze by adding gates as the monkeys' training progressed.

Monkeys moved their cursor, the solid red circle, by manipulating their joysticks. The cursor was free to move left or right, but once the monkey moved the cursor essentially into the first gate, the cursor's leftward movement was blocked. Now it could not move back out to decline the trial, and the monkey was committed to completing the (hopefully) Possible trial. Holding the joystick in the "Right" position produced a rightward movement of the cursor, which automatically adjusted along the vertical axis so as to allow the cursor to pass through the center of each gate.

Contacting the blue goal circle resulted in the blanking of the screen as a whoop sound played over speakers and the food hopper dispensed a pellet. A new trial was generated and displayed after 1 s. However, if the cursor contacted both circles of a gate (i.e., entering the constriction point in an Impossible trial), the control of the cursor was frozen for 4 s, leaving the screen visible for that time so that the monkey could fully perceive the constriction point that was the focus of the error. Then the screen went blank, a buzz sound played over speakers, and a trial-less time-out period ensued. Early in training, this time-out lasted 20 s. Later in training a time-out of 6 s was sufficient to sustain the monkeys' strong performance. The next trial followed. Contacting the trial-decline icon "?" resulted in a blank screen for 1 s, followed by a newly generated trial, and this next trial was chosen at random from the full range of trial types. That is, the use of the trial-decline response had no tangible benefit in changing the difficulty of the subsequent trial.

Training. The monkeys were trained in this task by making responses to Possible or Impossible trials with increasing numbers of

⁴ A standard subject analysis comparing latencies when constrictions appeared in the first half versus the last half of the wickets was also significant, $t(27) = 3.985$, $p < .001$, $d = .625$, as was a GLM using constriction position as the independent variable, $F(25, 525) = 4.247$, $p < .001$, $\eta_p^2 = .168$.

gates. We began with a single gate, to instill the idea that passage through the gate was the relevant physical feature of the task.

Results

One-gate training. In the earliest phase of training, we presented one wicket gate only. Only that wicket, the red cursor to be moved, and the blue goal circle were on the screen. The wicket was either constricted or generously open. Monkeys just needed to make this visual discrimination, which would then serve them well going forward as the mazes got longer.

In this phase, Obi completed 2,323 trials, 1,147 Possible trials and 1,176 Impossible trials. On Possible trials, he only declined 10 trials, or 1%. On Impossible trials, he declined 1,173 trials, or >99%. He made only three commission errors by launching into a constricted gate.

Lou completed 1,316 trials, 650 Possible trials and 666 Impossible trials. On Possible trials, he declined seven trials (1%). He made the appropriate decision to commit to approach the goal circle on 99% of trials. On Impossible trials, he declined 649 trials (97.5%). He made only 17 commission errors. Given one wicket gate, the monkeys had extremely sharp discriminations.

The monkeys were subsequently moved on to perform with maze displays of increasing visual complexity and spatial extent. During this progression, Obi performed with two-gate mazes (6,419 trials, 93% correct responses), four-gate mazes (1,703 trials, 83% correct), eight-gate mazes (3,434 trials, 68% correct), and 12-gate mazes (34,732 trials, 95% correct). Likewise, Lou performed with two-gate mazes (4,645 trials, 92% correct responses), four-gate mazes (2,260 trials, 87% correct), eight-gate mazes (9,956 trials, 79% correct), and 12-gate mazes (32,459 trials, 86% correct). As visual complexity and spatial extent increased, the monkeys' performance declined up through eight-gate mazes. However, then, especially given extensive 12-gate training, their discriminations recovered.

16-Gate testing. The monkeys now entered our stages of mature testing. At the stage of 16 gates, the maze displays were spatially quite extended, perhaps requiring a directional scanning strategy. The monkeys were also skilled and trained by now. Accordingly, we now made a more detailed examination of the monkeys' performance, including latency analyses to look for any systematic scanning strategies they might be using. In this 16-gate condition, we removed from the task the 1-s delay period imposed on performance to block impulsive responding. Now their cursor would move at any point after the trial illuminated. (We will compare their performance in this condition with their performance at 20 gates when the 1-s delay was reimposed).

In 16-gate training, Obi completed 7,610 trials, 3,778 Possible trials and 3,832 Impossible trials with one of the 16 gates constricted. On Possible trials, he declined 170 trials, or 4%. On Impossible trials, he declined 3,502 trials, or 91%. He was 22 times as likely to decline Impossible trials as Possible trials. He committed to 330 Impossible trials, or 9%. Overall, he made the appropriate response choice on 93% of trials.

Figure 3A shows Obi's serial-position curve for latency across gates. The monkeys' latencies were trimmed to exclude any latency beyond 5 s, a procedure like that we used with humans (a 15-s exclusion). But monkeys typically responded quickly, and so the lowered threshold is appropriate. (Having trimmed the RT

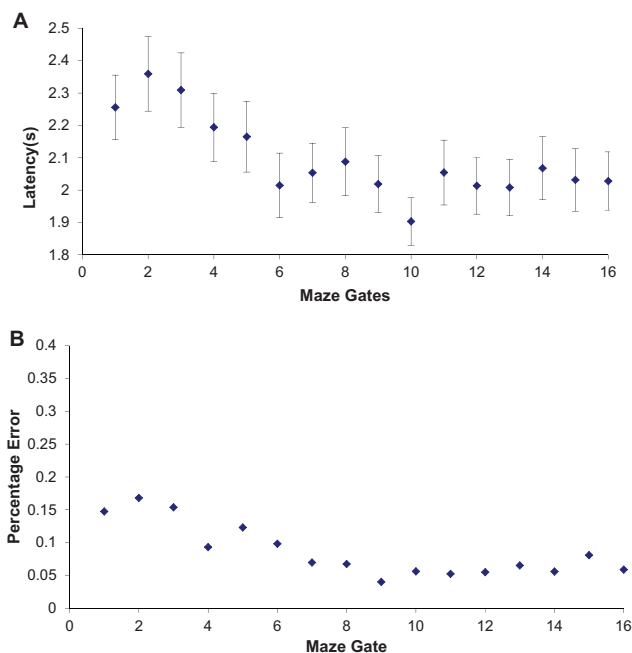


Figure 3. (A) Monkey Obi's mean latency for trial-decline responses on Impossible maze displays in Experiment 2 (16-gate condition). Each data point is framed by its 95% confidence interval. (B) Obi's percentage error when the constriction lay at each gate along the run of the maze. See the online article for the color version of this figure.

distributions to exclude outliers, the data pattern remained similar on using median instead of mean latencies. Across the 16 gate positions, mean and median latencies differed by 0.05 s on average.) With this trimming accomplished—260 latencies disqualified, or 7.4% of the 3,503 relevant trials—we analyzed 3,243 latency events. Obi's trial-decline responses were fastest when the gate constriction lay in the gates nearest the blue goal circle. He found those constrictions quicker. This strongly suggests that Obi scanned from right to left along the maze display, backward from the blue goal circle to his controllable red cursor. This is a completely appropriate strategy. Indeed, if Obi construed the problem as evaluating whether reaching the goal is possible, a search beginning at the most important goal spot could even be intuitive.

We confirmed this latency pattern as we did for humans. We compared the 1,030 trial-decline latencies observed when the constriction was in Gates 12–16 with the 956 latencies when the constriction was in Gates 1–5. These latencies were 2.03 s, 95% CI [1.99, 2.07], and 2.25 s, 95% CI [2.20, 2.30], respectively, standard deviations of .668 and .771, $t(1984) = 6.88$, $p < .001$, $d = .310$.

Here we add one descriptive result. Figure 3B shows Obi's percentage errors on Gates 1–16 of the maze. His best detection performance was nearest the blue goal circle, also suggesting a right-to-left directionality to his scan. Apparently, as Obi scanned more gates, he undercompensated for the need to spend more time and his detection sensitivity fell off as well.

In 16-gate testing, Lou completed 9,407 trials, 4,710 Possible trials and 4,697 Impossible trials. On Possible trials, he declined 1,148 trials, or 24%. On Impossible trials, he declined 3,879 trials, or 83%. He was three times as likely to decline Impossible trials as

Possible trials. He committed to 818 Impossible trials, or 17%. Across both trial types, he made the appropriate response choice on 79% of trials.

Lou's trial latencies were trimmed as already described, so that 318 latencies, or 8.2% of the 3,879 relevant trials, were disqualified, leaving 3,561 latency events analyzed. Figure 4A shows Lou's serial-position curve for latency across gates, calculated as before. (Here, too, the data pattern remained similar if we used median instead of mean latencies. Across the 16 gate positions, mean and median latencies only differed by 0.11 s on average.) Lou's fastest trial-decline latencies on Impossible trials fell at Gates 5–7. His maze-evaluation strategy may somehow have contained both directionalities, outward from this central place. We confirmed this latency pattern as follows. We compared the 403 trial-decline latencies observed when the constriction was in Gates 1 and 2 with the 782 latencies observed for Gates 5–7: means of 1.86 s, 95% CI [1.77, 1.96], and 1.72 s, 95% CI [1.66, 1.79], respectively, standard deviations of .968 and .923, $t(1183) = 2.33$, $p = .020$, $d = .144$. Likewise, we compared the 345 trial-decline latencies observed when the constriction was in Gates 15 and 16 with the 782 latencies observed for Gates 5–7: means of 1.98 s, 95% CI [1.89, 2.07], and 1.72 s, 95% CI [1.66, 1.79], respectively, standard deviations of .846 and .923, $t(1125) = 4.60$, $p < .001$, $d = .287$.

Figure 4B shows Lou's percentage errors on Gates 1–16 of the maze. As with Obi, his best detection performance tracked his fastest scanning time. Lou also undercompensated for the need to spend more time on gates late in his search process. Both monkeys

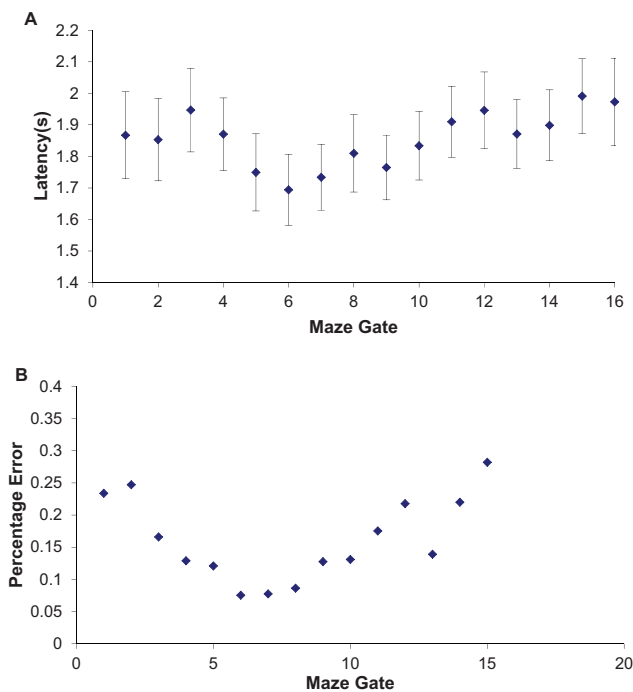


Figure 4. (A) Monkey Lou's mean latency for trial-decline responses on Impossible maze displays in Experiment 2 (16-gate condition). Each data point is framed by its 95% confidence interval. (B) Lou's percentage error when the constriction lay at each gate along the run of the maze. See the online article for the color version of this figure.

showed this general convergence between latency and error graphs, and we will not devote multiple figures to the error graphs later in the text. Interestingly, humans do not show this convergence. Their scanning phenomenon shows up only in the latency curves, because they better compensate for the need to give every wicket gate appropriate inspection. Perhaps for this reason, humans scan more slowly than monkeys.

20-Gate testing. In 20-gate testing, Obi completed 5,070 trials, 2,559 Possible trials and 2,511 Impossible trials with one of the 20 gates constricted. On Possible trials, he declined just 74 trials, or 3%. On Impossible trials, he declined 2,378 trials, or 95%. He was 30 times more likely to decline Impossible trials than Possible trials. He committed to just 133 Impossible trials. Across both trial types, he made the appropriate response choice on 96% of trials.

Figure 5A shows Obi's serial-position curve for latency across gates. Still his trial-decline responses were fastest when the gate constriction lay in the gates nearest the blue goal circle. Again, this suggests that he mainly began his search toward the right near the goal, and worked backward toward the left to where his response cursor lay. Considering testing on 16 gates and 20 gates combined, Obi showed this pattern over 12,680 trials.

We confirmed this latency pattern as before. We compared the 534 trial-decline latencies observed when the constriction was in Gates 1–5 with the 584 latencies when the constriction was in Gates 16–20. These latencies were 2.03 s, 95% CI [1.96, 2.10], and 1.68 s, 95% CI [1.62, 1.73], respectively, standard deviations of .827 and .646, $t(1116) = 7.85$, $p < .001$, $d = .472$.

In 20-gate testing, Lou completed 7,118 trials, 3,522 Possible trials and 3,596 Impossible trials. On Possible trials, he declined 622 trials, or 18%. On Impossible trials, he declined 2,737 trials, or 76%. He was four times more likely to decline Impossible trials than Possible trials. Across both trial types, he made the appropriate response choice on 79% of trials.

Figure 5B shows Lou's serial-position curve for latency across gates. Again, it suggested a scanning strategy from the middle out. Considering testing on 16 gates and 20 gates combined, Lou showed a similar performance pattern over 16,522 trials.

We compared the 265 trial-decline latencies observed when the constriction was in Gates 1 and 2 with the 449 latencies observed for Gates 5–7: means of 1.60 s, 95% CI [1.50, 1.70], and 1.51 s, 95% CI [1.43, 1.58], respectively, standard deviations of .826 and .842, $t(712) = 1.38$, $p = .168$, $d = .107$. We could not confirm a latency increase for the early gates—in this case Lou appeared to be more a left-to-right scanner. Likewise, we compared the 191 trial-decline latencies observed for constrictions in Gates 19 and 20 with the 449 latencies observed for constrictions in Gates 5–7: means of 1.75 s, 95% CI [1.64, 1.87], and 1.51 s, 95% CI [1.43, 1.58], respectively, standard deviations of .801 and .842, $t(638) = 3.52$, $p < .001$, $d = .301$.

Readers can compare Figures 5A and 3A for Obi, and Figures 5B and 4A for Lou, to consider a matter of interest to some. In the 20-gate condition, we tested the monkeys' response to having a 1-s delay imposed on their performance, so that the cursor would not move for the first second after the trial illuminated. We did so to see whether we could constructively block impulsive responding. Remember that in the 16-gate condition, this reflection period was not present. Our research explores the emergence of reflective mind in the primates, and we always seek manipulations that may

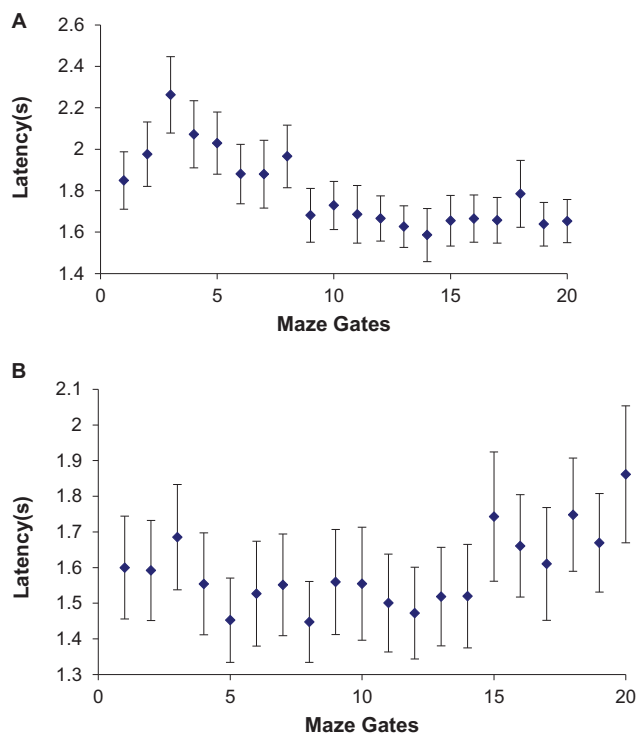


Figure 5. Obi's (A) and Lou's (B) mean latency for trial-decline responses on Impossible maze displays in Experiment 2 (20-gate condition). Each data point is framed by its 95% confidence interval. See the online article for the color version of this figure.

foster their highest potential. Comparing the two figures, though, one sees that our reflection manipulation did not change the character of the result. It does make the latencies in Figure 5 appear shorter than those in Figures 3 and 4, but, as was the case with humans in Experiment 1B, this arises because the monkeys got some of their scanning done before their trial clock initiated.

26-Gate testing. In 26-gate testing, Obi completed 10,059 trials, 4,974 Possible trials and 5,085 Impossible trials. On Possible trials, he declined 294 trials, or 6%. On Impossible trials, he declined 4,411 trials, or 87%. He was 14 times more likely to decline Impossible trials than Possible trials. He committed to just 674 Impossible trials. Across both trial types, he made the appropriate response on 90% of trials.

Figure 6A shows Obi's serial-position curve for latency. We confirmed the pattern of RTs as before. We compared the 744 trial-decline latencies observed when the constriction was in Gates 1–5 with the 762 latencies when the constriction was in Gates 22–26. This conservative analysis used the possible late-gate increase in latencies against our hypothesis test. These latencies were 2.38 s, 95% CI [2.31, 2.44], and 2.17 s, 95% CI [2.11, 2.22], respectively, standard deviations of .851 and .762, $t(1504) = 4.97$, $p < .001$, $d = .256$.

In 26-gate testing, Lou completed 7,072 trials, 3,435 Possible trials and 3,637 Impossible trials. On Possible trials, he declined 543 trials, or 16%. On Impossible trials, he declined 2,201 trials, or 61%. He was four times more likely to decline Impossible trials than Possible trials. Across both trial types, he made the appropriate response on 72% of trials.

Figure 6B shows Lou's serial-position curve for latency across gates. Now he barely showed an increase in scanning time to detect more rightward constrictions. We compared the 516 trial-decline latencies observed for constrictions in Gates 1–5 with the 222 latencies observed for Gates 22–26: means of 2.03 s, 95% CI [1.96, 2.11], and 2.24 s, 95% CI [2.13, 2.34], respectively, standard deviations of .875 and .807, $t(736) = 3.05$, $p = .002$, $d = .241$. Lou gave the impression in this condition that he was near his limit.

General Discussion

Summary

In four experiments, we asked humans to commit to completing spatially extended mazes or to decline them if they deemed the maze impossible. Half the mazes could be completed. Half contained a constriction in one gate that blocked completion. The optimal approach was to preevaluate the maze by scanning along the maze's length. We hoped that RT analyses would suggest that these scans occurred, if we observed that humans' latencies to declare mazes Impossible grew longer as the constriction point lay farther along the maze. Humans showed this scanning strategy. They scanned the mazes left to right, or right to left when the task

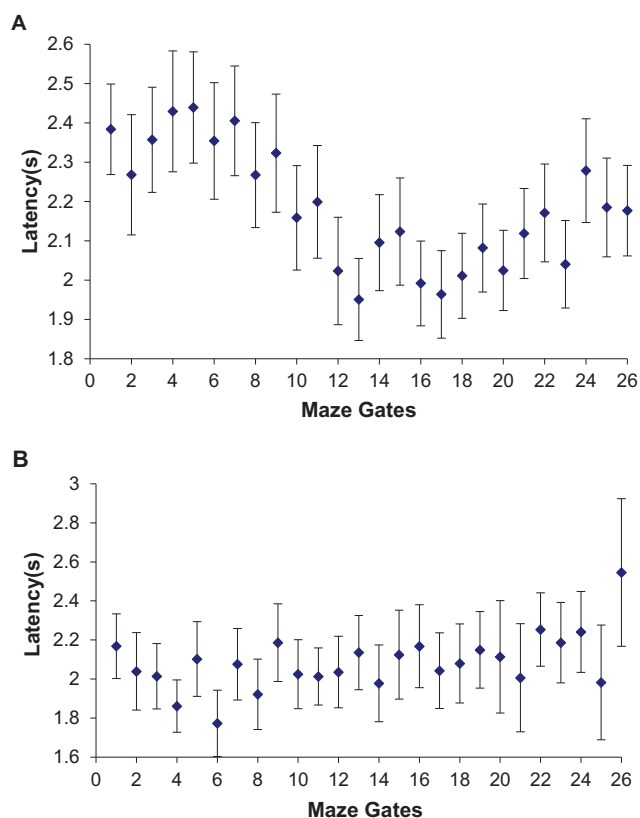


Figure 6. Obi's (A) and Lou's (B) mean latency for trial-decline responses on Impossible maze displays in Experiment 2 (26-gate condition). Each data point is framed by its 95% confidence interval. See the online article for the color version of this figure.

was mirror reflected, in the direction their cursor would move when they completed the maze.

Two macaques in the same paradigm also showed a systematic, preevaluation scanning strategy. One monkey apparently scanned beginning with the goal circle and working back toward his response cursor. One's scan may have begun internal to the maze. We discuss the general result—that monkeys, like humans, can complete a systematic scan of an extended visual array before deciding to decline Impossible trials—from two theoretical perspectives.

Results in the Context of Animal-Metacognition Research

Animals' trial-decline responses facing difficult trials have motivated an important comparative literature. However, the likelihood of accompanying associative-learning processes has had a strong impact on this literature, with high-level and low-level behavioral interpretations vying to explain the empirical findings. In fact, the field's inaugural demonstrations presented static and immediately present perceptual stimuli to animals (tones, clip arts, etc.), and animals declined trials in which difficult or ambiguous stimuli were presented that could cause error (but which had also caused past errors!). That last consideration is important. If animals avoid touching or responding to a difficult/ambiguous stimulus, this could be a matter of stimulus aversion/avoidance directed away from error-causing and reward-reducing stimuli. It might not represent a higher level process of difficulty monitoring at all. This has been the theoretical line taken by associative theorists toward many animal-metacognition findings.

Therefore, researchers have tried to dissociate animals' metacognitive performances from underlying associative-learning processes. Many researchers have taken on this challenge.

Researchers have shown the flexible transfer of the trial-decline response to new tasks and situations—sometimes even on the first trial of tasks (Kornell et al., 2007; Washburn, Smith, & Shields, 2006). Instantaneous transfer would hardly be expected by associative accounts in which learning would depend on training and associative learning in the new trial-specific contexts.

Researchers have shown that animals can metacognitively multitask, using the trial-decline response adaptively to manage difficulty in several different task contexts that are interleaved randomly trial by trial (Smith et al., 2010). This suggests that trial-decline responses are made on the basis of some more generalized assessment of difficulty or uncertainty that transcends stimulus-specific associations.

Researchers have shown that trial-decline responses may be especially resource intensive, particularly dependent on the cognitive resources available in working memory. For example, Smith, Coutinho, et al. (2013) showed that monkeys' primary perceptual responses in discrimination tasks were not impaired by the imposition of a concurrent working memory load. But their trial-decline responses were impaired. Likewise, humans' metacognitive judgments can be stifled by a working memory load under some circumstances (Coutinho et al., 2015; Schwartz, 2008). These findings suggest that trial-decline responses are representative of higher level cognitive processes to which low-level associative descriptions may not apply.

Researchers have shown that animals can instantaneously adjust their response strategy for managing difficult and uncertain trials, by becoming more error averse when they have more to lose. Associatively entrained response-strength gradients would not be expected to have this moment-to-moment changeability and flexibility (Zakrzewski et al., 2014).

Our research has a family resemblance to these studies. Here, we spatially extended our visual displays, thereby temporally extending the preevaluation process that was necessary to apprehend them entirely and respond adaptively. Macaques, like humans, apparently completed the temporally extended evaluation process before they settled on a response. This allowed them to sensitively discriminate between Possible and Impossible mazes. Their performance, and the temporal maps of their latencies, strongly suggests this systematic strategy.

Thus, this project joins its peers in broadening the empirical base showing sophisticated forms of difficulty monitoring in animals. It solidifies in the literature the growing theoretical consensus that for some animals the trial-decline response is part of a behavioral system that has some high-level cognitive elements and perhaps some precursor elements of metacognitive awareness. However, neither the present research nor its peer studies decisively prove that animals possess a close analog to humans' florid, conscious, verbal/declarative metacognition. In addition, neither the present research nor its peer studies deny or disprove that animals' performances in this area have important associative-learning underpinnings as well.

Results in the Context of Research on Animal Cognition

In our earliest animal studies, we made an intriguing observation. Sometimes on difficult trials, animals would almost move their joystick-controlled cursor to give a primary perceptual response, but then they would balk at the last instant and move the cursor to decline the trial. It was as though they had gotten cold feet. Sometimes on difficult trials, they would almost decline the trial, but then determinedly reverse course to choose a primary perceptual response. It was as though they were saying, "Wait, I know this!" The cursor's trajectory seemed to chart the animals' changes of mind. We joked—but in earnest—that if cursor movements had been prevalent in comparative psychology's early history and not bar presses, behaviorism never could have taken hold, because cursor movements show so clearly these changes of mind.

Tolman (1932/1967, 1938)—much earlier—raised the same issues when he saw rats dithering and vacillating at the choice point of a T-maze. These vacillating movements were his infamous virtual trial and error movements. He thought that virtual trial and error movements could reveal rats' online cognitive processes. Even more strikingly, he thought they could become *the behaviorist's definition of animal consciousness* (Tolman, 1927).

We believe that the ability to see and measure the course and trajectory and directionality of an animal's cognitive processes is an intriguing, profoundly important possibility. Yet there are still few systematic explorations of animal minds that try to map these time courses and trajectories. For example, there is scant research assaying the time course of imagining, memory scanning, mental rotation, and so forth (but see Neiwirth & Rilling, 1987; Sands & Wright, 1982).

Illustrating this possibility, we explored animals' scanning of visual arrays spatially extended to the point that they could not be apprehended in one glance. We placed macaques into a task in which they needed some systematic way to preevaluate the whole array, meanwhile deferring response, until they had collected the required information. And we were able to chart the time course of their systematic preevaluation of Possible and Impossible mazes, studying in the clear an epoch of information processing during which they undertook a systematic, self-terminating search of the extended maze array. Of course, our paradigm is not the only one that one could bring to this demonstration. For one example, one might study monkeys' eye movements in our task. This would be an exciting, complementary approach toward revealing their systematic search strategies. For another example, one could consider other visual-search paradigms that have been used with monkeys (Ipata, Gee, Goldberg, & Bisley, 2006; Motter & Belky, 1998; Purcell et al., 2010). In these paradigms, animals are rewarded for visually fixating a stimulus target. These paradigms may well tap deliberate search processes like those we describe here—especially if the target is the conjunction of two visual features. However, often in these tasks the rapid search integrates seamlessly with the successful saccade, minimizing the appearance of the systematic search process and making its independent analysis difficult. A strong feature of our paradigm is that we made the arrays so spatially extended that the search process became temporally extended, systematic, easily observable, and potentially manipulable.

The present demonstration doesn't disprove associative learning, or prove metacognition, or downplay associative learning, or anything like that. Probably animals were searching for, and reacting to, narrow wickets as associative stimuli in our task. That doesn't affect the demonstration. The systematic nature of the cognitive search remains separate from that.

Comparative psychology has a deep bench of paradigms suitable for applying this kind of chrometric analysis and suitable for incorporating a trial-decline response. For one example, there is productive work on planning to perform sequences of response choices (Beran & Parrish, 2012; Biro & Matsuzawa, 1999). There are productive studies of maze performance and planning by nonhuman primates when the mazes are two-dimensional, of different complexities, and pose different levels of inhibitory challenge (perhaps requiring deviations away from the goal position at a crucial moment—Beran, Parrish, Futch, Evans, & Perdue, 2015; Fragaszy, Johnson-Pynn, Hirsh, & Brakke, 2003; Fragaszy et al., 2009; Mushiake, Saito, Sakamoto, Sato, & Tanji, 2001; Pan et al., 2011). It would be interesting to study the time course of primates' planning performances in these mazes, as well as other species such as pigeons (Miyata, Ushitani, Adachi, & Fujita, 2006), asking whether more complex mazes and more inhibition-challenging mazes take more extensive cognitive preplanning work and time. For this purpose, one would naturally borrow techniques from the present research—using half Impossible mazes, but also giving animals a trial-decline response with which to ward off Impossible trials.

The key element of this methodology is to temporally and spatially extend the stimulus display presented to animals, thereby segregating the stages of animals' information processing, to separate animals' inspection, reflection, and consideration processes from their eventual response actions (which may certainly some-

times involve responses to associative cues uncovered in the task). Given this separation, researchers can observe the animal's cognitive processes clearly and somewhat independently from their associative-learning processes. In this way, researchers may be able to assess more sensitively how fluently, or not, animals operate on a cognitive plane of information processing, and the strength of their cognitive continuities with humans.

References

- 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. <http://dx.doi.org/10.1111/j.1467-7687.2008.00725.x>
- Basile, B. M., & Hampton, R. R. (2014). Metacognition as discrimination: Commentary on Smith et al. (2014). *Journal of Comparative Psychology*, *128*, 135–137. <http://dx.doi.org/10.1037/a0034412>
- Basile, B. M., Schroeder, G. R., Brown, E. K., Templer, V. L., & Hampton, R. R. (2015). Evaluation of seven hypotheses for metamemory performance in rhesus monkeys. *Journal of Experimental Psychology: General*, *144*, 85–102. <http://dx.doi.org/10.1037/xge0000031>
- Beran, M. J., Evans, T. A., Klein, E. D., & Einstein, G. O. (2012). Rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*) remember future responses in a computerized task. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*, 233–243. <http://dx.doi.org/10.1037/a0027796>
- Beran, M. J., & Parrish, A. E. (2012). Sequential responding and planning in capuchin monkeys (*Cebus apella*). *Animal Cognition*, *15*, 1085–1094. <http://dx.doi.org/10.1007/s10071-012-0532-8>
- Beran, M. J., Parrish, A. E., Futch, S. E., Evans, T. A., & Perdue, B. M. (2015). Looking ahead? Computerized maze task performance by chimpanzees (*Pan troglodytes*), rhesus monkeys (*Macaca mulatta*), capuchin monkeys (*Cebus apella*), and human children (*Homo sapiens*). *Journal of Comparative Psychology*, *129*, 160–173. <http://dx.doi.org/10.1037/a0038936>
- Beran, M. J., Perdue, B. M., & Smith, J. D. (2014). What are my chances? Closing the gap in uncertainty monitoring between rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*, 303–316. <http://dx.doi.org/10.1037/xan0000020>
- Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Cognition*, *120*, 90–105. <http://dx.doi.org/10.1016/j.cognition.2011.02.016>
- Biro, D., & Matsuzawa, T. (1999). Numerical ordering in a chimpanzee (*Pan troglodytes*): Planning, executing, and monitoring. *Journal of Comparative Psychology*, *113*, 178–185. <http://dx.doi.org/10.1037/0735-7036.113.2.178>
- Boysen, S. T., & Berntson, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *103*, 23–31. <http://dx.doi.org/10.1037/0735-7036.103.1.23>
- Brown, A. L., Bransford, J. D., Ferrara, R. A., & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology* (Vol. 3, pp. 77–164). New York, NY: Wiley.
- Call, J. (2010). Do apes know that they could be wrong? *Animal Cognition*, *13*, 689–700. <http://dx.doi.org/10.1007/s10071-010-0317-x>
- Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language*, *23*, 58–89. <http://dx.doi.org/10.1111/j.1468-0017.2007.00329.x>
- Carruthers, P., & Ritchie, J. B. (2012). The emergence of metacognition: Affect and uncertainty in animals. In M. J. Beran, J. Brandl, J. Perner, & J. Proust (Eds.), *Foundations of metacognition* (pp. 76–93). Oxford, United Kingdom: Oxford University Press. <http://dx.doi.org/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. <http://dx.doi.org/10.1037/a0020129>
- Coutinho, M. V. C., Redford, J. S., Church, B. A., Zakrzewski, A. C., Couchman, J. J., & Smith, J. D. (2015). The interplay between uncertainty monitoring and working memory: Can metacognition become automatic? *Memory and Cognition, 43*, 990–1006. <http://dx.doi.org/10.3758/s13421-015-0527-1>
- Debert, P., Matos, M. A., & McIlvane, W. (2007). Conditional relations with compound abstract stimuli using a go/no-go procedure. *Journal of the Experimental Analysis of Behavior, 87*, 89–96. <http://dx.doi.org/10.1901/jeab.2007.46-05>
- Dunlosky, J., & Bjork, R. A. (Eds.). (2008). *Handbook of metamemory and memory*. New York, NY: Psychology Press.
- Dunlosky, J., & Metcalfe, J. (2008). *Metacognition*. New York, NY: Sage.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*, 906–911. <http://dx.doi.org/10.1037/0003-066X.34.10.906>
- Flemming, T. M., Beran, M. J., & Washburn, D. A. (2007). Disconnect in concept learning by rhesus monkeys (*Macaca mulatta*): Judgment of relations and relations-between-relations. *Journal of Experimental Psychology: Animal Behavior Processes, 33*, 55–63. <http://dx.doi.org/10.1037/0097-7403.33.1.55>
- Footo, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology, 17*, 551–555. <http://dx.doi.org/10.1016/j.cub.2007.01.061>
- Fragaszy, D., Johnson-Pynn, J., Hirsh, E., & Brakke, K. (2003). Strategic navigation of two-dimensional alley mazes: Comparing capuchin monkeys and chimpanzees. *Animal Cognition, 6*, 149–160. <http://dx.doi.org/10.1007/s10071-002-0137-8>
- Fragaszy, D. M., Kennedy, E., Murnane, A., Menzel, C., Brewer, G., Johnson-Pynn, J., & Hopkins, W. (2009). Navigating two-dimensional mazes: Chimpanzees (*Pan troglodytes*) and capuchins (*Cebus apella* sp.) profit from experience differently. *Animal Cognition, 12*, 491–504. <http://dx.doi.org/10.1007/s10071-008-0210-z>
- Fujita, K. (2009). Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition, 12*, 575–585. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/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. <http://dx.doi.org/10.3819/ccbr.2009.40002>
- Herman, L. M., & Forestell, P. H. (1985). Reporting presence or absence of named objects by a language-trained dolphin. *Neuroscience and Biobehavioral Reviews, 9*, 667–681. [http://dx.doi.org/10.1016/0149-7634\(85\)90013-2](http://dx.doi.org/10.1016/0149-7634(85)90013-2)
- Ipata, A. E., Gee, A. L., Goldberg, M. E., & Bisley, J. W. (2006). Activity in the lateral intraparietal area predicts the goal and latency of saccades in a free-viewing visual search task. *The Journal of Neuroscience, 26*, 3656–3661. <http://dx.doi.org/10.1523/JNEUROSCI.5074-05.2006>
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. T. (2009a). Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews, 4*, 29–39. <http://dx.doi.org/10.3819/ccbr.2009.40003>
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. T. (2009b). Reinforcement and metacognition. *Comparative Cognition and Behavior Reviews, 4*, 58–60. <http://dx.doi.org/10.3819/ccbr.2009.40007>
- Katz, J. S., Wright, A. A., & Bachevalier, J. (2002). Mechanisms of same/different abstract-concept learning by rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes, 28*, 358–368. <http://dx.doi.org/10.1037/0097-7403.28.4.358>
- Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289–326). Cambridge, United Kingdom: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9780511816789.012>
- 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*, 297–315. <http://dx.doi.org/10.1037/0096-3445.123.3.297>
- Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Psychological Science, 18*, 11–15. <http://dx.doi.org/10.1111/j.1467-8721.2009.01597.x>
- Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*, 64–71. <http://dx.doi.org/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. <http://dx.doi.org/10.1037/a0026478>
- Le Pelley, M. E. (2014). Primate polemic: Commentary on Smith, Couchman, and Beran (2014). *Journal of Comparative Psychology, 128*, 132–134. <http://dx.doi.org/10.1037/a0034227>
- Metcalfe, J. (2008). Evolution of metacognition. In J. Dunlosky & R. A. Bjork (Eds.), *Handbook of metamemory and memory* (pp. 29–46). New York, NY: Psychology Press.
- Metcalfe, 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. <http://dx.doi.org/10.1093/acprof:oso/9780195161564.003.0002>
- Miyata, H., Ushitani, T., Adachi, I., & Fujita, K. (2006). Performance of pigeons (*Columba livia*) on maze problems presented on the LCD screen: In search for preplanning ability in an avian species. *Journal of Comparative Psychology, 120*, 358–366. <http://dx.doi.org/10.1037/0735-7036.120.4.358>
- Morgan, C. L. (1906). *An introduction to comparative psychology*. London, United Kingdom: Walter Scott.
- Motter, B. C., & Belky, E. J. (1998). The guidance of eye movements during active visual search. *Vision Research, 38*, 1805–1815. [http://dx.doi.org/10.1016/S0042-6989\(97\)00349-0](http://dx.doi.org/10.1016/S0042-6989(97)00349-0)
- Mushiaki, H., Saito, N., Sakamoto, K., Sato, Y., & Tanji, J. (2001). Visually based path-planning by Japanese monkeys. *Cognitive Brain Research, 11*, 165–169. [http://dx.doi.org/10.1016/S0926-6410\(00\)00067-7](http://dx.doi.org/10.1016/S0926-6410(00)00067-7)
- Neiworth, J. J., & Rilling, M. E. (1987). A method for studying imagery in animals. *Journal of Experimental Psychology: Animal Behavior Processes, 13*, 203–214. <http://dx.doi.org/10.1037/0097-7403.13.3.203>
- Nelson, T. O. (1992). *Metacognition: Core readings*. Toronto, Canada: Allyn & Bacon.
- Nelson, T. O. (1996). Consciousness and metacognition. *American Psychologist, 51*, 102–116. <http://dx.doi.org/10.1037/0003-066X.51.2.102>
- Pan, J., Kennedy, E. H., Pickering, T., Menzel, C. R., Stone, B. W., & Fragaszy, D. M. (2011). Development of maze navigation by tufted capuchins (*Cebus apella*). *Behavioural Processes, 86*, 206–215. <http://dx.doi.org/10.1016/j.beproc.2010.11.006>
- 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. <http://dx.doi.org/10.1007/s10071-005-0007-2>

- Paul, E. J., Smith, J. D., Valentin, V. V., Turner, B. O., Barbey, A. K., & Ashby, F. G. (2015). Neural networks underlying the metacognitive uncertainty response. *Cortex: A Journal Devoted to the Study of the Nervous System and Behavior*, *71*, 306–322. <http://dx.doi.org/10.1016/j.cortex.2015.07.028>
- Pepperberg, I. M. (1983). Cognition in the African Grey parrot: Preliminary evidence for auditory/vocal comprehension of the class concept. *Animal Learning and Behavior*, *11*, 179–185. <http://dx.doi.org/10.3758/BF03199646>
- Premack, D. (1978). On the abstractness of human concepts: Why it would be difficult to talk to a pigeon. In S. H. Hulse, H. Fowler, & W. K. Honig (Eds.), *Cognitive processes in animal behavior* (pp. 423–451). Hillsdale, NJ: Erlbaum.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*, 515–526. <http://dx.doi.org/10.1017/S0140525X00076512>
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113–1143. <http://dx.doi.org/10.1037/a0020311>
- Roberts, W. A., Feeney, M. C., McMillan, N., MacPherson, K., Musolino, E., & Petter, M. (2009). Do pigeons (*Columba livia*) study for a test? *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 129–142. <http://dx.doi.org/10.1037/a0013722>
- Sands, S. F., & Wright, A. A. (1982). Monkey and human pictorial memory scanning. *Science*, *216*, 1333–1334. <http://dx.doi.org/10.1126/science.7079768>
- Savage-Rumbaugh, E. S. (1986). *Ape language: From conditioned response to symbol*. New York, NY: Columbia University Press.
- Schwartz, B. L. (2008). Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory and Cognition*, *36*, 9–19. <http://dx.doi.org/10.3758/MC.36.1.9>
- 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. <http://dx.doi.org/10.1037/0096-3445.126.2.147>
- Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, *13*, 389–396. <http://dx.doi.org/10.1016/j.tics.2009.06.009>
- Smith, J. D., Beran, M. J., & Couchman, J. J. (2012). Animal metacognition. In E. A. Wasserman & T. R. Zentall (Eds.), *The Oxford handbook of comparative cognition* (2nd ed., pp. 282–304). Oxford, United Kingdom: Oxford University Press.
- Smith, J. D., Beran, M. J., Couchman, J. J., & Coutinho, M. V. C. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin and Review*, *15*, 679–691. <http://dx.doi.org/10.3758/PBR.15.4.679>
- 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 of London Series B, Biological Sciences*, *367*, 1297–1309. <http://dx.doi.org/10.1098/rstb.2011.0366>
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2014a). Animal metacognition: A tale of two comparative psychologies. *Journal of Comparative Psychology*, *128*, 115–131. <http://dx.doi.org/10.1037/a0033105>
- Smith, J. D., Couchman, J. J., & Beran, M. J. (2014b). A tale of two comparative psychologies: Reply to commentaries. *Journal of Comparative Psychology*, *128*, 140–142. <http://dx.doi.org/10.1037/a0034784>
- 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. <http://dx.doi.org/10.1037/a0029601>
- Smith, J. D., Flemming, T. M., Boomer, J., Beran, M. J., & Church, B. A. (2013). Fading perceptual resemblance: A path for macaques to conceptual matching? *Cognition*, *129*, 598–614. <http://dx.doi.org/10.1016/j.cognition.2013.08.001>
- Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal Cognition*, *13*, 93–101. <http://dx.doi.org/10.1007/s10071-009-0249-5>
- 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. <http://dx.doi.org/10.1037/0096-3445.124.4.391>
- Smith, J. D., Shields, W. E., Schull, J., & Washburn, D. A. (1997). The uncertain response in humans and animals. *Cognition*, *62*, 75–97. [http://dx.doi.org/10.1016/S0010-0277\(96\)00726-3](http://dx.doi.org/10.1016/S0010-0277(96)00726-3)
- Staddon, J. E. R., Jozefowicz, J., & Cerutti, D. T. (2007, April 13). Metacognition: A problem not a process. *PsyCrit*, 1–5. Retrieved from http://psycrit.com/w/Metacognition:_A_problem_not_a_process
- Suda-King, C. (2008). Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition*, *11*, 21–42. <http://dx.doi.org/10.1007/s10071-007-0082-7>
- 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. <http://dx.doi.org/10.1037/0097-7403.34.2.266>
- 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. <http://dx.doi.org/10.1007/s10071-011-0468-4>
- Tolman, E. C. (1927). A behaviorist's definition of consciousness. *Psychological Review*, *34*, 433–439. <http://dx.doi.org/10.1037/h0072254>
- Tolman, E. C. (1967). *Purposive behavior in animals and men*. New York, NY: Appleton-Century-Croft. (Original work published 1932)
- Tolman, E. C. (1938). The determiners of behavior at a choice point. *Psychological Review*, *45*, 1–41. <http://dx.doi.org/10.1037/h0062733>
- 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. <http://dx.doi.org/10.1098/rsbl.2009.0737>
- 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*, 157–164. <http://dx.doi.org/10.3758/BF03203490>
- 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. <http://dx.doi.org/10.1037/0097-7403.32.2.185>
- Zakrzewski, A. C., Perdue, B. M., Beran, M. J., Church, B. A., & Smith, J. D. (2014). Cashing out: The decisional flexibility of uncertainty responses in rhesus macaques (*Macaca mulatta*) and humans (*Homo sapiens*). *Journal of Experimental Psychology: Animal Learning and Cognition*, *40*, 490–501. <http://dx.doi.org/10.1037/xan0000041>

Received April 9, 2017

Revision received October 23, 2017

Accepted October 31, 2017 ■