Psychological Topics, 34 (2025), 1, 97-115

Original Scientific Paper UDC: 159.929.072
591.5
https://doi.org/10.31820/pt.34.1.5 599.8

Will Monkeys Wager Differently as a Function of Stimulus Fluency or When Making Immediate Versus Delayed Judgments of Memory?

Michael J. Beran and Kathleen Friedlein

Georgia State University, Department of Psychology and Language Research Center

Abstract

Nonhuman animals can engage in forms of metacognitive control and monitoring processes. However, very little testing of the relation between fluency and metacognition has been done in animals, and little research has assessed memory performance in relation to animals making immediate versus delayed judgments of their memory. Here, wagers made by monkeys during test trials served as a form of confidence measure of how well they could complete a memory test. These wagers occurred either after the delay interval between the sample presentation and the test (delayed judgments) or after the sample presentation but before the delay interval and the test (immediate judgments). Overall, no significant difference in performance was found between these two conditions. We also manipulated the fluency of stimuli by either contrasting small (low fluency) or large (high fluency) stimuli or by manipulating size and the degree to which stimuli were of similar perceptual classes (low fluency, harder to distinguish stimuli such as triangular shapes) or were dissimilar in color and shape (high fluency, clip art images). Although low fluency stimuli were remembered at lower levels, the monkeys showed no evidence of adjusting wagering behavior as a function of stimulus type. Thus, the present experiment showed no evidence that monkeys benefitted from delay of judgments of memory and no evidence of stimulus fluency affecting their confidence as measured by their wagering. Rather, most monkeys preferred consistent wagers across all trial types. This may indicate a metacognitive limitation or some other form of behavioral satisficing that led to suboptimal performance.

Keywords: metacognition, monkeys, judgments of memory, prospection, stimulus fluency, delayed matching to sample

---

Michael J. Beran https://orcid.org/0000-0002-2504-0338

This project was supported by funding from the National Science Foundation through award BCS2043667.

Acknowledgment

The authors thank Brooke Jackson, Andres Sanchez, Liz Haseltine, Joseph McKeon, Robert Vera, Katarina Syed, and Dara Jonkoski for assistance with data collection.

✉ Michael J. Beran, Department of Psychology and Language Research Center, Georgia State University, P.O. Box 5010, Atlanta, GA 30302-5010 USA. E-mail: *mberan1@gsu.edu*

Introduction

Adult humans engage in a sophisticated form of thinking known as *metacognition.* They monitor their perceptual experiences and conceptual knowledge states, they assess the availability and strength of specific memories, and they anticipate their capacity to respond appropriately and adaptively to new cognitive challenges (Benjamin et al., 1998; Dunlosky & Metcalfe, 2009; Flavell, 1979; Metcalfe & Shimamura, 1994; Nelson, 1992; Schwartz, 1994; Schwartz & Metcalfe, 1994). Metacognition is highly relevant to educational achievement (e.g., learning strategies, test-taking skills; Dunlosky & Connor, 1997; Roebers, 2002; Schneider, 2008), and it is often held to be a central aspect of higher intelligence in humans.

There has been a longstanding debate about whether other species can engage in metacognition. For 25 years, empirical data have been generated from numerous species to aid in this debate, and to provide a comparative perspective on metacognition (Basile et al., 2015; Beran et al., 2012; Brown et al., 2017; Call, 2010; Carruthers, 2008; Crystal, 2014; Crystal & Foote, 2009; Hampton, 2009; Kornell, 2009, 2014; Smith, 2009, 2010; Templer & Hampton, 2012). There is consensus that at least some species show patterns of performance in various tasks that require interpretations based on executive functioning and cognitive control, if not on metacognition fully (e.g., Brown et al., 2017, 2019; Carruthers, 2014; Smith et al., 2013; Templer & Hampton, 2012; Zakrzewski et al., 2014). Some of these performances of animals are not reducible to Stimulus-Response associative learning, reward histories, or specifically trained stimulus classes and tasks (e.g., Basile et al., 2015; Beran et al., 2013; Brown et al., 2017; Kornell et al., 2007; Morgan et al., 2014; Smith et al., 2006, 2008)

The field now needs to document and understand how internal cues trigger monitoring processes that afford metacognitive judgments in nonhuman animals. Cues such as stimulus fluency might influence metacognitive decisions across species in ways similar to how we know such cues affect humans. For example, the ease with which a stimulus is viewed, such as its larger size or more vivid color, can affect people's metacognitive judgments of how well they can remember or categorize such stimuli (e.g., Mueller et al., 2014; Murphy et al., 2022; Rhodes & Castel, 2008; Reber & Schwarz, 1999; Susser et al., 2013; Thompson et al., 2013; Undorf & Zimdahl, 2019; Undorf et al., 2017; for review, see Alter & Oppenheimer, 2009). As one example, Rhodes and Castel (2008) showed that words presented in larger font sizes during a serial recall task were judged to be better learned than those in smaller fonts, even though performance during recall was not affected by this manipulation. This is a form of metacognitive illusion, and an effect that has rarely been assessed in nonhuman species. Ferrigno et al. (2017) is an exception to this as they showed that rhesus macaques would make metacognitive errors based on stimulus fluency. Specifically, monkeys were more willing to make higher confidence wagers on their memory accuracy based on higher perceptual fluency of to-be-remembered stimuli, even though fluency itself was not related to objective performance on the memory test.

Another important function of metacognition is judging when learning is sufficient to proceed to taking a test or when learners wager or otherwise report their confidence in what they have learned (i.e., judgments of learning, or JOLs). The *delayed-JOL effect* shows that JOLs are more accurate for later recall when they are made a short time after study rather than immediately following study (e.g., Dunlosky & Nelson, 1992). Different theoretical mechanisms for the delayed-JOL effect have been offered, including that delaying these judgments may focus attention solely on long-term memory versus also having access to short-term memory or that encoding and recall contexts are slightly more different after a delay (Nelson & Dunlosky, 1991). Other ideas have been offered, including that memory strength stochastically drifts across the retention interval (Sikström & Jönsson, 2005). JOLs might also serve as additional encoding opportunities, and if so, then delayed judgments would lead to better encoding of items that were remembered (e.g., Kimball & Metcalfe, 2003). Comparative data on whether this effect occurs would help distinguish mechanisms reliant on language or other uniquely human capacities. Although animals cannot provide specific numerical or verbal assessments of how well they have learned, they can be trained to use wagers or other behavioral indicators that could reflect confidence in learning, or confidence in memory as in the case of monkeys in the Ferrigno et al. (2017) study. This idea forms the basis for the present experiment.

This experiment combines parts of the memory testing design used by Ferrigno et al. (2017) with a manipulation of the delayed JOL effect. Specifically, we adapted the prospective condition they used in which wagers were made after memory encoding but prior to completing the memory assessment (in contrast to retrospective wagers that were made after the memory test was completed). In our task, monkeys engaged in a computerized delayed matching-to-sample task. They saw a sample image which then disappeared, and it was either of high perceptual fluency (i.e., brightly colored, large, and easy to distinguish from distractor stimuli) or of low fluency (i.e., small and difficult to recognize among distractor stimuli). After a delay period during which the monkeys had to remember the stimulus, they had to find that stimulus from an array of four options. They also provided a confidence wager on their ability to complete the trial correctly. High wagers led to four rewards when correct, but a 30-second timeout when incorrect. Low wagers led to two rewards when correct and a 10-second timeout when incorrect. This design was previously successfully used by monkeys in our lab (Shields et al., 2005), and it conceptually matched the wagering approach of Ferrigno et al. (2017). To investigate the delayed JOL effect, we included two test conditions that varied when the JOL was made. In the *immediate JOL condition*, this wager was made as soon as the image left the screen (i.e., at the start of the delay interval). In the *delayed JOL condition*, this wager occurred 10 s after the image left the screen, right before the delay interval ended. Thus, we could assess performance relative to wagering to test whether JOLs were made more accurately when they were delayed. Note that in this experiment, we are designating JOLs as assessments of how likely one will be able to remember (immediate) or how well one has remembered (delayed) a sample image that was encoded. We recognize that this is not the same as many human JOL experimental paradigms, as those studies also can include retrospective JOLs and other forms of commenting on confidence, but it is an approach that can provide initial data on whether nonhuman animals may be capable of more sophisticated forms of judgments of learning.

The rationale for our design was to assess how stimulus fluency might influence (i.e., modulate) wagering behavior and, specifically, whether monkeys might show evidence of a metacognitive error. Such errors have been proposed as an essential next step in better understanding the cues and mechanisms underlying “meta”-level representations in animals (Kornell, 2014). In our design, metacognitive errors would be evident if monkeys’ wagering was higher when stimulus fluency was higher (i.e., when stimuli were more fluent) even though actual memory performance was not equally affected by high fluency stimuli compared to lower fluency stimuli. An alternative outcome would be that manipulations to fluency would not affect wagering and would only affect performance. This would suggest that, at least in this task, metacognitive errors do not occur, presumably because monkeys respond based solely on memory trace strength or perhaps even through a simple wagering strategy.

Methods

Participants

We tested 17 capuchin monkeys and five rhesus macaques (see Table 1). All monkeys were housed at the Language Research Center of Georgia State University. All monkeys had previous exposure to delayed matching to sample tasks and were familiar with the testing apparatus. Monkeys were never food or water-deprived, and testing was always voluntary. Monkeys also had access to indoor and outdoor facilities and enrichment items outside of their individual testing boxes. The monkeys were socially housed for the majority of the day with conspecifics (capuchin monkeys) or were paired daily with a socially compatible partner with whom they could spend time indoors and outdoors (macaques). The only exception was macaque Lou who was under medical treatment during the experiment that required single housing. He still had visual and auditory access to conspecifics at all times. All experimental procedures were approved by the Institutional Animal Care and Use Committee of Georgia State University. Georgia State University is accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International.

Table 1

Participant Demographic Table and Training Performance

Species	Sex	Age / Birth Year	Sessions	Trials to criterion
Rhesus Monkeys				
Chewie	M	24 / 1999	15	1,689
Han	M	21 / 2002	13	492
Lou	M	30 / 1993	13	4,261
Luke	M	24 / 1999	10	1,020
Murph	M	30 / 1993	5	1,048
Capuchin Monkeys				
Albert	M	12 / 2011	22	614
Applesauce	F	19 / 2004	17	3,947
Atilla*	M	11 / 2012	38	5,582
Bailey*	F	24 / 1999	15	2,401
Bias*	F	36 / 1987	22	651
Gambit	F	27 / 1996	9	486
Gretel	F	20 / 2003	11	623
Griffin	M	26 / 1997	11	586
Ingrid	F	11 / 2012	16	3,817
Ivory	F	25 / 1998	6	1,636
Liam	M	20 / 2003	7	2,648
Lily	F	26 / 1997	13	2,656
Logan	M	18 / 2005	7	409
Lychee*	F	24 / 1999	23	2,654
Nala	F	21 / 2002	11	1,151
Nkima	M	16 / 2007	22	1,177
Wren	F	21 / 2002	8	596

* Failed to meet criterion to advance to the testing phases.

Apparatus

Monkeys completed test sessions using the Language Research Center testing system. This included a personal computer, a 17-inch monitor display, a manipulable joystick controller, and a pellet dispenser. Monkeys were separated from the monitor by a clear face plate and were able to manipulate the joystick and access pellet rewards with their hands through cutouts in the face plate. When a correct answer was selected, rewards were dispensed from the pellet dispenser through a plastic tube and into the monkey's testing box. Rewards were 45 mg banana flavored pellets (bio-serv.com).

General Task

Across all phases of the experiment, at the start of each trial, a red cursor, 4.5 mm in diameter, was located at the bottom center of a white screen and a grey rectangle, 24 mm tall and 60 mm wide, was placed at the mid-top center of the screen. The monkey controlled the cursor by moving a joystick with its hand, and joystick deflections led to smooth movement of the cursor on the screen. When the monkey moved the cursor into contact with the grey rectangle, the rectangle disappeared, and a sample image appeared onscreen. This image differed across phases of the experiment (see below), and it was present for one second before disappearing from the screen. This was the to-be-remembered sample, and its size varied across different phases of the experiment (see below). The monkey's task was to choose the image that looked the same from four options (each 60 mm x 60 mm) presented on screen later in the trial. Chance performance was 25% given these four choices. Those choice options appeared in the four corners of the screen, and contacting the item that was identical to the sample led to 2 or 4 food pellets (condition dependent, see below) being delivered, after which the screen was cleared of all images. If the correct item was selected, the trial initiation stimulus was presented. If an incorrect image was selected, the screen was cleared and a 10- or 30-second (condition-dependent) timeout period occurred, during which the screen remained blank. No food reward was delivered. Next, the trial initiation stimulus was presented.

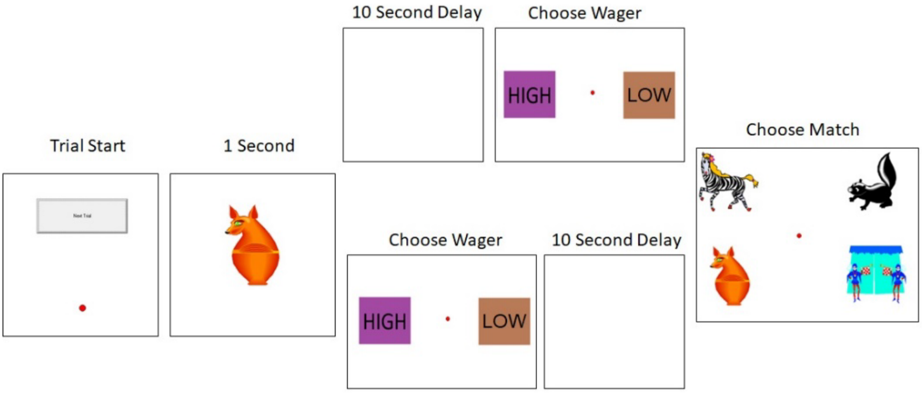
Training Phase

In this phase, sample images were comprised of colored clip art images ($N = 350$) that were all high in perceptual fluency in terms of their size (60 mm x 60 mm), coloration, and distinctiveness. After contacting the sample image, one of two types of trials was presented. In the *immediate wagering condition*, monkeys were shown a wager option immediately after seeing the to-be-remembered stimulus. After wager selection, monkeys experienced a 10-second delay before the memory test. In the *delayed wagering condition*, monkeys experienced a 10-second delay before they made a wagering decision. In this condition, the memory test was taken immediately after wager selection (see Figure 1).

The monkeys were presented with only one wagering option on each trial, so that they could learn the contingencies for each option. These options were presented as two unique stimuli: a purple box (48 mm high and 60 mm wide) with the word HIGH written in black letters on the mid-left side of the screen, or a brown box with the word LOW written in yellow letters on the mid-right side of the screen. Selection of a high wager led to four food rewards when monkeys correctly identified the to-be-remembered stimulus, and a 30-second time out when an incorrect choice was made, during which the screen was cleared and remained blank. Selection of the low wager led to two food rewards upon correct image identification, and a 10-second screen time out when an incorrect selection was made. Our initial criterion for passing this phase was a score of 50% or higher accuracy on each of the possible training conditions (immediate low wagers, immediate high wagers, delayed low wagers, delayed high wagers). Criterion was assessed at the conclusion of each test

Figure 1

Schematic of the Trial Types



Note. High wagers led to 4 pellets if correctly completed and a 30-second timeout if incorrect, and Low wagers led to 2 pellets if correct and a 10-second timeout if incorrect.

session once a monkey had completed at least 400 total trials. Four capuchin monkeys failed to meet this criterion. After 1,427 trials, Attila was still at chance levels (~25% correct) in all conditions. Bias was not willing to engage in trials in this condition (she only completed 651 trials across 22 test sessions). Bailey (2,401 trials) and Lychee (2,654 trials) had one or more conditions in which they did not meet criterion. One capuchin monkey (Liam) had one of the four conditions in which he did not quite reach the 50% criterion (48.7% correct), but he did meet criterion for the other conditions, so we chose to progress him to keep the sample size as large as possible.

Test Phase 1

A total of 13 capuchin monkeys and 5 rhesus macaques met the criteria to enter the testing phase. In this phase, 50% of the trials were immediate wagering trials, and 50% were delayed wagering trials. On each trial, there was a 50% probability that both wagering options were offered and a 50% probability of a “forced” trial, where only one wagering option was offered (as in the Training phase). This was done for two reasons. First, to ensure that monkeys had continual experience with each of the wagers in terms of their contingencies, and second, to provide interleaved data on performance when wagering was not controlled by the monkeys compared to when it was. When both choices were available for wagering, they were in the same locations as they had been in the Training phase, but both were available simultaneously. All other aspects of the trials were the same as in the Training phase. Monkeys completed 2,000 trials in this phase.

Test Phase 2

In this phase, each monkey again completed 2,000 trials. The change to the design involved the to-be-remembered stimuli. Our first manipulation of perceptual fluency involved stimulus size. On each trial, the sample was randomly determined to have high or low perceptual fluency. High fluency images were defined as being large (60 mm x 60 mm), and low fluency images were defined as being small (48 mm x 48 mm). In each of these conditions, the match choices were of the same size as the sample on each trial (large or small). All other aspects of the trials were consistent with the previous test phase.

For reasons unrelated to this study, two capuchin monkeys (Applesauce and Albert) were discontinued from the study.

Test Phase 3

In the last test phase, the perceptual fluency and distinctiveness of the to-be-remembered stimulus were manipulated. This design change was in response to extreme biases seen during Test Phase 2 in an effort to reduce the proportion of high wagers being made by so many of the monkeys (see Results). Stimuli continued to be presented as either small or large, but we also varied the type of stimulus presented. More fluent images remained the same clip art images that had been used previously, whereas low fluency images all consisted of abstract shapes. These shapes all took the form of red-colored pointed shapes presented against a black background. There were 150 of these stimuli that could be presented along with the 350 clip art images. On trials with clip art samples, only clip art match choices were presented, and the same was true for the shape stimuli. In addition, within those 150 shapes, we used 10 distinct classes of shapes, so that the most difficult trials could include one or more incorrect options that came from the same class of shape (see Appendix 1 for all of these stimuli). Stimuli were drawn at random, meaning that there was approximately a 30% probability of a trial involving these shape stimuli as the sample and a 70% probability of a clip art stimulus being the same. All other aspects of trial presentation remained consistent with the previous test phase. Each monkey again completed 2,000 trials, except for Lily who completed 1,876 due to computer error. Due to low participation rates (i.e., refusing to enter the text boxes for testing sessions), one capuchin monkey (Ivory) was discontinued from the study.

Results

Training Phase

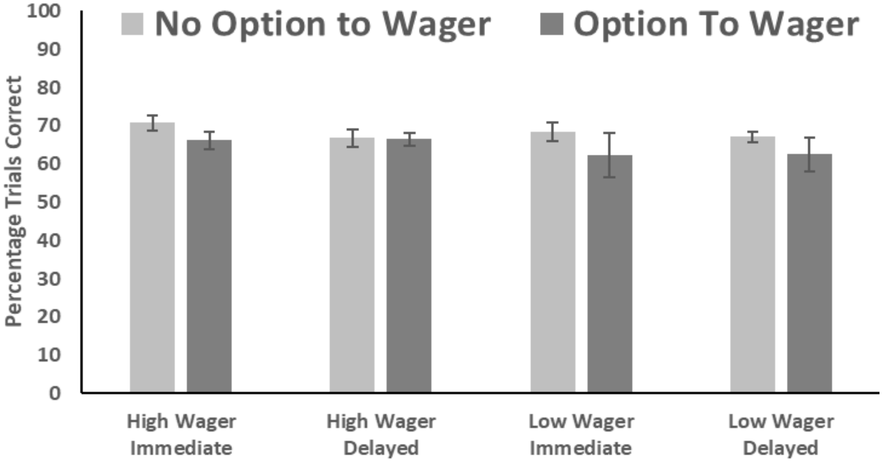
Table 1 presents the number of sessions and the numbers of completed trials by all monkeys in this phase. As noted above, four monkeys did not progress to the testing phases.

Test Phase 1

We first examined performance (percentage of trials correct in matching) on all possible trial types. A mixed-factors ANOVA was conducted with species as a between-subjects factor, wagering judgment condition (immediate or delayed), trial type (forced or choice wagering), and chosen wager amount (high or low wager) as within-subjects factors. The only significant effect was for trial type (forced or choice wagering), $F(1, 16) = 7.05$, $p = .017$, $\eta p^2 = .31$. All other main effects, and all interactions were non-significant, all $p > .05$. For the effect of trial type, performance was significantly better when monkeys were forced to choose the only available wager than when they could choose a wager. The data combined across species in each condition, and each wager period and type of wager are shown in Figure 2.

Figure 2

Performance of all Monkeys in Testing Phase 1.



Note. Error bars depict corrected 95% confidence intervals for repeated measures tests (Cousineau, 2005).

We also assessed whether, on choice trials where either wager could be made, monkeys preferred to wager high more often in the immediate or delayed judgment conditions. Here, we used another mixed ANOVA with species as the between-subjects factor and wagering judgment condition (immediate or delayed wager) as the within-subjects factor. There was no effect of species, *F*(1, 16) = 0.07, *p* = .80, ηp^2 = .004, no effect of wagering judgment condition, *F*(1, 16) = 2.87, *p* = .11, ηp^2 = .15., and no interaction, *F*(1, 16) = 0.01, *p* = .97, ηp^2 = .00. In general, the group of monkeys preferred to make high wagers overall for each wagering judgment condition; immediate condition: one sample *t*(17) = 2.55, *p* = .021, delayed condition: one-sample *t*(17) = 3.55, *p* = .002.

Test Phase 2

In Test Phase 2, extreme biases emerged for most of the monkeys' wager choices (Table 2). Eleven monkeys (seven capuchin monkeys and four rhesus macaques) chose the high wager on 90% or more of all trials. Three other capuchin monkeys instead chose the low wager on 90% or more of trials in these four conditions. Only two monkeys chose the high wager and low wager at rates that suggested some discernment across trials about that decision. Macaque Chewie selected the high wager on 75% of trials, and capuchin Nala selected it on 58% of trials. Given these extreme biases, it was clear that the monkeys did not provide data relevant to our main experimental question.

Table 2

Percentages of Choices of the High Wager for all Monkeys in Test Phase 2 and Test Phase 3

Species	Test Phase 2				Test Phase 3			
	Large		Small		Large		Small	
	Imm.	Delay	Imm.	Delay	Imm.	Delay	Imm.	Delay
Rhesus Monkeys								
Chewie	78.6	71.5	73.9	78.3	96.1	97.6	96.4	97.1
Han	97.2	94.8	96.6	94.3	99.2	98.4	98.8	98.3
Lou	99.6	100.0	100.0	99.2	97.2	97.7	97.3	96.4
Luke	98.1	99.6	98.7	99.2	97.6	98.5	99.6	98.0
Murph	93.3	98.4	95.0	99.6	97.3	99.2	99.0	98.3
Capuchin Monkeys								
Gambit	0.39	0.72	0.84	0.37	2.3	1.1	2.1	2.0
Gretel	99.4	99.1	97.1	98.8	99.0	100.0	99.2	100.0
Griffin	98.4	98.8	98.0	99.2	92.1	89.5	92.0	89.3
Ingrid	99.6	100.0	100.0	100.0	99.6	100	99.2	100.0
Ivory	11.9	5.0	14.1	7.1	NA	NA	NA	NA
Liam	87.1	95.1	100	99.4	100	100	100	100.0
Lily	98.0	99.2	97.8	98.6	95.9	99.6	96.3	99.5
Logan	100.0	99.3	100.0	100.0	100.0	100.0	100.0	100.0
Nala	21.1	96.0	18.6	93.9	47.5	98.5	37.1	99.2
Nkima	0.4	0	0	0.4	2.1	6.8	5.1	8.2
Wren	97.0	95.7	96.2	97.7	97.2	100	98.0	99.6

To attempt to understand these strong biases, we examined performance levels on trials where the wager was forced, so that we had enough data to assess both species. Performance was significantly better when stimuli were large (72.2%) compared to when they were small (68.6%), *t*(14) = 2.46, *p* = .028. This confirmed that our fluency measure did generate performance changes, although these were not substantial. Performance was also better when immediate wagers were made (80.0%), compared to delayed wagers (68.9%), *t*(14) = 2.35, *p* = .035. This pattern is less informative because on these trials the wager was not under control, and so it is not clear why this would be true except that perhaps knowing the wager ahead of the delay increased the monkeys' attention to their memory because there was no competing need to choose between wager options. Finally, we examined overall performance (correctness) for trials where a wager had to be selected versus trials where the wager was forced, to see whether performance again differed as it had in Test Phase 1. We did not find that effect in this phase, with performance in the wagering condition (72.3% correct) and forced condition (71.0% correct) not significantly different from each other, paired *t*(15) = .47, *p* = .64.

Test Phase 3

The monkeys again showed extreme biases in their wager choices, and in the same direction as for Phase 2 (Table 2). Twelve monkeys (seven capuchin monkeys and five rhesus macaques) chose the high wager on 90% or more of all trials. Two capuchins instead chose the low wager on 90% or more trials. Only one monkey (Nala) chose the high wager and low wager at rates that suggested some discernment across trials about that decision. She selected the high wager on 71% of trials.

Again, to investigate whether these biases were suboptimal from the perspective of the difficulty of different trial types, we examined performance as a function of stimulus type, stimulus size, and the immediate or delayed wager that was to be made. We collapsed across all choice trials and forced trials to examine these effects for all data. This 3-way ANOVA indicated that there was a main effect of stimulus type, *F*(1, 14) = 417.9, *p* < .001, with performance significantly better on trials with clip art stimuli (*M* = 75.9%, *SD* = 8.9) than triangle shapes (*M* = 28.4%, *SD* = 7.1). There was no main effect of stimulus size, *F*(1, 14) = 0.19, *p* = .67, and there was no main effect of the immediate or delayed trial type, *F*(1, 14) = 0.29, *p* = .64. There was a significant interaction of stimulus type and stimulus size, *F*(1, 14) = 10.66, *p* = .006. For clip art stimuli, larger stimuli were remembered at a higher level than smaller stimuli, *t*(14) = 2.50, *p* = .026, but this was not the case for the pointed triangular stimuli, *t*(14) = -1.21, *p* = .25.

Thus, as in Phase 2, our manipulations of stimuli did generate performance differences, and in this case, for stimulus type, it was a substantial difference, with monkeys performing three times higher with clip art stimuli than pointed triangles. However, the monkeys did not modify their wagering choices at all relative to Phase 2.

Finally, we again examined overall performance (correctness) for trials where a wager had to be selected versus trials where the wager was forced, to test whether performance differed as it had in Test Phase 1. We did not find that effect in this phase, with performance in the wagering condition (65.2% correct) and forced condition (65.2% correct) not significantly different from each other, paired $t(14) = .0032$, $p = .98$.

Discussion

Monkeys consistently failed to show evidence of adjusting their wagering behavior based on the manipulations we employed in this experiment. In the first phase, we simply presented immediate or delayed wagering opportunities to assess whether wagers differed. This would have indicated that the monkeys had some sense of different likelihood of success in the task at each wager point. However, there was no indication of differential wagering in this phase.

In Phase 2, we varied the size of the sample image, and this was to provide a simple method of increasing stimulus salience on this single property. Monkeys' wagering behavior did not vary as a function of this, and it was possible that this was not a sufficiently noticeable difference in the stimuli that had to be remembered even though the monkeys performed better on the task with the larger stimuli.

In Phase 3, we used stimuli that gave monkeys either easier-to-remember (clip art) or harder-to-remember (shapes) stimuli. This manipulation was reflected objectively in performance levels. However, once again, the monkeys failed to differentially wager as a function of stimulus difficulty, and also again as a function of stimulus size. In essence, the monkeys were not connecting wagers to likelihood of success in this phase, and that was true regardless of whether wagers were made immediately or after a delay.

The clear conclusion is that, in this experiment, monkeys showed no evidence of a delayed judgment of learning effect, and no evidence of even monitoring memory or anticipating memory performance at all. Of course, this does not mean monkeys cannot monitor and control their memory, as such demonstrations have occurred in a variety of other metamemory tasks (e.g., Basile et al., 2015; Brown et al., 2019; Hampton, 2001; Hampton et al., 2004; Kornell et al., 2007; Morgan et al., 2014; Tanaka & Funahashi, 2012; Templer & Hampton, 2012; Templer et al., 2018). What these results mean is that, in this experiment, no such evidence was present.

That leaves us with the question of what to make of these largely *null* results. One possibility is that the monkeys were not motivated to perform more effectively, or to adjust their wagering. We think this is unlikely, especially for Phase 3, where errors occurred often with the red shapes, and monkeys sat through a great duration of timeouts. Timeouts of this duration are typically highly effective in producing learning in many other tasks we have designed, including uncertainty monitoring tasks where hard stimuli can be avoided through an escape response (e.g., Smith et al., 2006, 2013). Thus, this does not seem likely, but we should acknowledge that the use of different reward amounts (two and four pellets) may have been highly salient to the monkeys and more so than other task characteristics.

Another possibility is that the monkeys did not understand the nature of the wagering responses, or even the contingencies of those responses. However, our use of forced trials throughout the experiment is the best approach we have to at least ensure they did experience the contingencies of wagering high and wagering low after making each of those wager types. Across training and three phases of the experiment, such forced "reminder" trials did not induce changes in wagering in the free choice trials. When given the choice, monkeys nearly all showed strong and near-total biases, and for most monkeys, this was for the high wager.

Although some past work has shown that monkeys can make wagers on their performance (Ferrigno et al., 2017; Shields et al., 2005), our experience is that such performances are much harder to elicit compared to other kinds of metacognitive responses. The present *null* results mimic another series of past experiments in which we found that capuchin monkeys were highly reluctant to make escape responses in psychophysical discrimination tasks until such time as that chance responding was so low that those responses were extremely adaptive to make. This contrasted with most macaque monkeys who were more likely to engage with escape responses and other metacognitive responses across a variety of experiments (e.g., Beran & Smith, 2011; Beran et al., 2014; Perdue et al., 2015; Smith et al., 2018). The present experiment did not produce any species differences, but it may be that the reward structure and the 25% chance level of responding were sufficient to allow monkeys to show biases to only one wager value. The question is why this would be true.

Our preferred explanation is that in this task the wagering component of the design may have simply been one for which the monkeys did not have sufficient cognitive resources to engage such wagering proficiently. Remembering a stimulus and making an intervening wager may have been too taxing to that ongoing rehearsal. This was true even though our task, at least with clip art images, allowed for the use of stimulus familiarity to correctly complete trials, and so working memory should not have been taxed (see Basile & Hampton, 2013; Brady & Hampton, 2018; Brown & Hampton, 2020). Perhaps making the wager response was still too distracting or interfering for the monkeys to engage in a decisional process rather than a rote (biased) response. If true, this highlights a limitation of our design, in the sense that other approaches may allow wagers or other forms of judging one's learning or memory to not interfere or distract as much from the learning and remembering itself. This will need to be investigated in future variations of this kind of task. Our data only supported this possibility in Test Phase 1, where trials with forced wagers overall were performed at a statistically significantly higher level, although this reflected only a few percentage points of difference. This was not a difference seen in Test Phase 2 or Test Phase 3, so the present results are ambiguous as to this issue.

At present, we conclude that rhesus monkeys and capuchin monkeys do not flexibly wager on how likely they will remember a stimulus (when judgments are immediate) or how well they have remembered a stimulus (when judgments are delayed). This is despite sufficient experience in how wagers work in terms of payoffs and penalties, and with sufficiently difficult stimuli that pushed performance to low levels compared to performance with much easier stimuli. We conclude that this is a metacognitive limitation of these species, at least in this type of task, and one that might reflect cognitive overloading of memory and decisional processes in the general cognition of these species. However, that conclusion requires additional testing using other types of wagers or judgments of learning and memory to ensure the present results are not an artefact of our design.

References

Alter, A. L., & Oppenheimer, D. M. (2009). Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review, 13*(3), 219–235. https://doi.org/10.1177/1088868309341564

Basile, B. M., & Hampton, R. R. (2013). Dissociation of active working memory and passive recognition in rhesus monkeys. *Cognition, 126*(3), 391–396. https://doi.org/10.1016/j.cognition.2012.10.012

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*(1), 85–102. https://doi.org/10.1037/xge0000031

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. https://doi.org/10.1037/0096-3445.127.1.55

Beran, M. J., Brandl, J., Perner, J., & Proust, J. (Eds.) (2012). *Foundations of metacognition*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199646739.001.0001

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*(3), 303–316. https://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*(1), 90–105. https://doi.org/10.1016/j.cognition.2011.02.016

Beran, M. J., Smith, J. D., & Perdue, B. M. (2013). Language-trained chimpanzees name what they have seen, but look first at what they have not seen. *Psychological Science, 24,* 660–666. https://doi.org/10.1177/0956797612458936

Brady, R. J., & Hampton, R. R. (2018). Nonverbal working memory for novel images in rhesus monkeys. *Current Biology, 28*(24), 3903–3910. https://doi.org/10.1016/j.cub.2018.10.025

Brown, E. K., Basile, B. M., Templer, V. L., & Hampton, R. R. (2019). Dissociation of memory signals for metamemory in rhesus monkeys *(Macaca mulatta*). *Animal Cognition, 22,* 331–341. https://doi.org/10.1007/s10071-019-01246-5

Brown, E. K., & Hampton, R. R. (2020). Cognitive control of working memory but not familiarity in rhesus monkeys (*Macaca mulatta*). *Learning & Behavior, 48*, 444–452. https://doi.org/10.3758/s13420-020-00432-7

Brown, E. K., Templer, V. L., & Hampton, R. R. (2017). An assessment of domain-general metacognitive responding in rhesus monkeys. *Behavioural Processes, 135,* 132–144. https://doi.org/10.1016/j.beproc.2016.12.004

Call J. (2010). Do apes know that they can be wrong? *Animal Cognition, 13,* 689–700. https://doi.org/10.1007/s10071-010-0317-x

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind and Language, 23,* 58–89. https://doi.org/10.1111/j.1468-0017.2007.00329.x

Carruthers, P. (2014). Two concepts of metacognition. *Journal of Comparative Psychology, 128,* 138–139. https://doi.org/10.1037/a0033877

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042

Crystal, J. D. (2014). Where is the skepticism in animal metacognition? *Journal of Comparative Psychology, 128,* 152–154. https://doi.org/10.1037/a0034427

Crystal, J. D., & Foote, A. L. (2009). Metacognition in animals: Trends and challenges. *Comparative Cognition and Behavior Reviews, 4,* 54–55. https://doi.org/10.3819/ccbr.2009.40005

Dunlosky, J., & Connor, L. T. (1997). Age differences in the allocation of study time account for age differences in memory performance. *Memory & Cognition, 25,* 691–700. https://doi.org/10.3758/BF03211311

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition*. Sage Publications.

Dunlosky, J., & Nelson, T. O. (1992). Importance of the kind of cue for judgments of learning (JOL) and the delayed-JOL effect. *Memory & Cognition, 20,* 374–380. https://doi.org/10.3758/BF03210921

Ferrigno, S., Kornell, N., & Cantlon, J. F. (2017). A metacognitive illusion in monkeys. *Proceedings of the Royal Society B: Biological Sciences, 284*(1862), Article 20171541. https://doi.org/10.1098/rspb.2017.1541

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist, 34*(10), 906–911. https://doi.org/10.1037/0003-066X.34.10.906

Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences, 98*(9), 5359–5362. https://doi.org/10.1073/pnas.071600998

Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition & Behavior Reviews, 4*, 17–28. https://doi.org/10.3819/ccbr.2009.40002

Hampton, R. R., Zivin, A., & Murray, E. A. (2004). Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition, 7,* 239–246. https://doi.org/10.1007/s10071-004-0215-1

Kimball, D. R., & Metcalfe, J. (2003). Delaying judgments of learning affects memory, not metamemory. *Memory & Cognition, 31,* 918–929. https://doi.org/10.3758/bf03196445

Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Cognitive Science, 18,* 11–15. https://doi.org/10.1111/j.1467-8721.2009.01597.x

Kornell, N. (2014). Where is the “meta” in animal metacognition? *Journal of Comparative Psychology, 128,* 143–149. https://doi.org/10.1037/a0033444

Kornell, N., Son, L. K., & Terrace, H. S. (2007). Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science, 18*(1), 64–71. https://doi.org/10.1111/j.1467-9280.2007.01850.x

Metcalfe, J., & Shimamura, A. P. (Eds.). (1994). *Metacognition: Knowing about knowing*. MIT Press. https://doi.org/10.7551/mitpress/4561.001.0001

Morgan, G., Kornell, N., Kornblum, T., & Terrace, H. S. (2014). Retrospective and prospective metacognitive judgments in rhesus macaques (*Macaca mulatta*). *Animal Cognition, 17,* 249–257. https://doi.org/10.1007/s10071-013-0657-4

Mueller, M. L., Dunlosky, J., Tauber, S. K., & Rhodes, M. G. (2014). The font-size effect on judgments of learning: Does it exemplify fluency effects or reflect people’s beliefs about memory? *Journal of Memory and Language, 70,* 1–12. https://doi.org/10.1016/j.jml.2013.09.007

Murphy, D. H., Huckins, S. C., Rhodes, M. G., & Castel, A. D. (2022). The effect of perceptual processing fluency and value on metacognition and remembering. *Psychonomic Bulletin & Review, 29*(3), 910–921. https://doi.org/10.3758/s13423-021-02030-8

Nelson, T. O. (Ed.) (1992). *Metacognition: Core readings.* Allyn and Bacon.

Nelson, T. O., & Dunlosky, J. (1991). When people’s judgments of learning (JOLs) are extremely accurate at predicting subsequent recall: The “delayed-JOL effect.” *Psychological Science, 2*(4), 267–271. https://doi.org/10.1111/j.1467-9280.1991.tb00147.x

Perdue, B. M., Church, B. A., Smith, J. D., & Beran, M. J. (2015). Exploring potential mechanisms underlying the lack of uncertainty monitoring in capuchin monkeys. *International Journal of Comparative Psychology*, *28*. https://doi.org/10.46867/ijcp.2015.28.00.05

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition, 8*(3), 338–342. https://doi.org/10.1006/ccog.1999.0386

Rhodes, M. G., & Castel, A. D. (2008). Memory predictions are influenced by perceptual information: Evidence for metacognitive illusions. *Journal of Experimental Psychology: General, 137*(4), 615–625. https://doi.org/10.1037/a0013684

Roebers, C. M. (2002). Confidence judgments in children's and adult's event recall and suggestibility. *Developmental Psychology, 38*(6), 1052–1067. https://doi.org/10.1037/0012-1649.38.6.1052

Schneider, W. (2008). The development of metacognitive knowledge in children and adolescents: Major trends and implications for education. *Mind, Brain, and Education, 2*(3), 114–121. https://doi.org/10.1111/j.1751-228X.2008.00041.x

Schwartz, B. L. (1994). Sources of information in metamemory: Judgments of learning and feelings of knowing. *Psychonomic Bulletin & Review, 1*(3), 357–375. https://doi.org/10.3758/BF03213977

Schwartz, B. L., & Metcalfe, J. (1994). Methodological problems and pitfalls in the study of human metacognition. *Metacognition: Knowing about Knowing*, 93–113. https://doi.org/10.7551/mitpress/4561.003.0007

Shields, W. E., Smith, J. D., Guttmannova, K., & Washburn, D. A. (2005). Confidence judgments by humans and rhesus monkeys. *The Journal of General Psychology, 132*(2), 165–186. https://pmc.ncbi.nlm.nih.gov/articles/PMC4607308/pdf/nihms14266.pdf

Sikström, S., & Jönsson, F. (2005). A model for stochastic drift in memory strength to account for judgments of learning. *Psychological Review, 112*(4), 932–950. https://doi.org/10.1037/0033-295X.112.4.932

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences, 13,* 389–396. https://doi.org/10.1016/j.tics.2009.06.009

Smith, J. D. (2010). Inaugurating the study of animal metacognition. *International Journal of Comparative Psychology, 23,* 401–413. https://doi.org/10.46867/IJCP.2010.23.03.02

Smith J. D., Beran, M. J., Coutinho, M. V. C., & Couchman, J. C. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin and Review, 15,* 679–691. https://doi.org/10.3758/PBR.15.4.679

Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General, 135,* 282–297. https://doi.org/10.1037/0096-3445.135.2.282

Smith, J. D., Coutinho, M. V. C., Church, B., & Beran, M. J. (2013). Executive-attentional uncertainty responses by rhesus monkeys (*Macaca mulatta*). *Journal of Experimental Psychology: General, 142,* 458–475. https://doi.org/10.1037/a0029601

Smith, T. R., Smith, J. D., & Beran, M. J. (2018). Not knowing what one knows: A meaningful failure of metacognition in capuchin monkeys. *Animal Behavior and Cognition, 5*(1), 55–67. https://doi.org/10.26451/abc.05.01.05.2018

Susser, J. A., Mulligan, N. W., & Besken, M. (2013). The effects of list composition and perceptual fluency on judgments of learning (JOLs). *Memory & Cognition, 41,* 1000–1011. https://doi.org/10.3758/s13421-013-0323-8

Tanaka, A., & Funahashi, S. (2012). Macaque monkeys exhibit behavioral signs of metamemory in an oculomotor working memory task. *Behavioural Brain Research, 233*(2), 256–270. https://doi.org/10.1016/j.bbr.2012.05.013

Templer, V. L., Brown, E. K., & Hampton, R. R. (2018). Rhesus monkeys metacognitively monitor memories of the order of events. *Scientific Reports, 8*(1), Article 11541. https://doi.org/10.1038/s41598-018-30001-y

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. https://doi.org/10.1007/s10071-011-0468-4

Thompson, V. A., Turner, J. A. P., Pennycook, G., Ball, L. J., Brack, H., Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237–251. https://doi.org/10.1016/j.cognition.2012.09.012

Undorf, M., & Zimdahl, M. F. (2019). Metamemory and memory for a wide range of font sizes: What is the contribution of perceptual fluency? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 45*(1), 97–109. https://doi.org/10.1037/xlm0000571

Undorf, M., Zimdahl, M. F., & Bernstein, D. M. (2017). Perceptual fluency contributes to effects of stimulus size on judgments of learning. *Journal of Memory and Language, 92,* 293–304. https://doi.org/10.1016/j.jml.2016.07.003

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. https://doi.org/10.1037/xan0000041

Received: October 16, 2024

Appendix

These were the difficult images used in Phase 3. Each of the first images in the leftmost column was randomly generated and then 14 additional, perceptually related images were generated from that prototype (the other 14 images in each row). Monkeys could not know before they wagered whether one, two, or three foil (incorrect) choices would be variations of the sample image or other red shapes, as this was randomly determined on each trial, the design of these stimuli ensured that this was a task with much higher difficulty than the trials with clip art images.