Distribution Agreement The text below should be reproduced exactly as written, on the Distribution Agreement page. Sign the page on the signature line, and type your name under the signature line. Write the date on the date line. **Distribution Agreement**

Signature:

______________________________ _______________

Emily Kathryn Brown Date

Rhesus monkeys generalize metacognitive responding across perceptual and memory tasks

By

Emily Kathryn Brown
MA
Psychology: Neuroscience and Animal Behavior

Robert Hampton
Advisor

Patricia Bauer
Committee Member

Joseph Manns
Committee Member

Donna Maney
Committee Member

Accepted:

Lisa A. Tedesco, Ph.D. Dean of the James T. Laney School of Graduate Studies

---

Date

Rhesus monkeys generalize metacognitive responding across perceptual and memory tasks

By

Emily Kathryn Brown
AB, Vassar College, 2006

Advisor: Robert R. Hampton, PhD

An abstract of
a thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in
Psychology
2015

Abstract
Rhesus monkeys generalize metacognitive responding across perceptual and memory tasks

By
Emily Kathryn Brown

Metacognition is the ability to access one's own cognitive states. To better understand the mechanisms that underlie apparently metacognitive behavior in nonhumans, we must determine what discriminative cues control decision making. Here, we evaluate the bases of metacognitive responding in rhesus monkeys (*Macaca mulatta*) by a series of generalization tests that varied in the types of cues available to control metacognition. In Experiment 1, we trained a group of monkeys on a perceptual discrimination in which correct responses were rewarded with food. We then provided a secondary metacognitive "*decline-test*" response by which monkeys could avoid tests and receive a guaranteed, but smaller, reward. In Experiments 2-4, we evaluated whether monkeys generalized use of the *decline-test* response to 3 new perceptual tasks, in the absence of stimulus-specific cues from their initial training. In Experiment 5, we further evaluated the possibility that metacognitive responding was controlled by public, non-introspective cues in a delayed matching-to-sample task. This memory test differed from the perceptual tests in that metacognitive responding cannot be controlled by properties of the test display. In Experiment 6, monkeys' use of the *decline-test* response differed as a function of task difficulty in a prospective judgment memory test. Because the metacognitive choice was presented before the test appeared in this final experiment, the public cues available to control metacognitive behavior were limited. Together these results provide provisional evidence that rhesus monkeys may be able to use a domain-general, private cue such as "uncertainty" to monitor that status of cognitive processes and knowledge states.

Rhesus monkeys generalize metacognitive responding across perceptual and memory tasks

By

Emily Kathryn Brown
AB, Vassar College, 2006

Advisor: Robert R. Hampton, PhD

A thesis submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Master of Arts
in
Psychology
2015

Acknowledgements

I thank my advisor, Robert R. Hampton, for valuable support and guidance. I thank Victoria L. Templer for discussions of the research question. I thank Dina P. Chou, Steven L. Sherrin, Jessica A. Joiner, and Tara Dove VanWormer for help running subjects. This work was supported by the National Science Foundation (grants IOS-1146316, BCS-0745573) and the National Institutes of Health (grant RO1MH082819).

Table of Contents

Figures and Tables

1. Introduction

Metacognition is the ability to access one's own cognitive states. To better understand the mechanisms that underlie apparently metacognitive behavior, we must determine what discriminative cues control decision making (Basile & Hampton, in press; Hampton, 2009a). These discriminative cues likely vary in the extent to which they are publically or privately available. Public cues depend on external information such as prior experience and reinforcement history. For example, if asked by a friend if I could spell a word, I would not need to access my private, cognitive state to accurately predict my ability to answer. Instead, I might base this judgment on my relative expertise with the subject matter—I am better at spelling than she is. Or, I could base my judgment on relevant prior experiences with those sorts of questions—I have been helpful with telling her how to spell words in the past. In the case of these public cues, a well-informed outsider could make a prediction about my performance equally reliable to my own prediction. In contrast, private cues depend on internal cognitive states, to which only the subject has special, privileged access. If my friend asked me to spell a familiar but challenging word, I might have to try to remember the spelling to gauge my ability to answer that particular question before I would know if I could help. An outsider could not accurately gauge information that I could only gain by introspection. Both public and private cues can likely produce behavioral modifications that are useful to learning (Hampton, 2009b). But humans are able to base metacognitive judgments on both public and private cues, whereas other animals may only have access to public cues (Kornell, 2013). Experimental dissociation of public and private cue use in nonhumans is important because it informs our understanding of the evolution of cognitive processes.

Comparative cognition could thus help to elucidate the relation, if one exists, between public and private cues. However, public and private cues are difficult to dissociate.

Researchers have developed paradigms to determine the bases of apparently metacognitive judgments in nonhumans. Subjects complete a primary task which varies in difficulty and is expected to elicit cognition. Subjects are then given the opportunity to make a secondary, metacognitive discrimination which allows them to selectively respond to avoid difficult trials on which they are unlikely to succeed. A major aim of metacognition research in nonhumans is to identify performance based on private cues, typically by ruling out use of public cues (Shettleworth & Sutton, 2003).

Many metacognition paradigms for nonhumans have used psychophysical discriminations as the primary task. The discriminanda differ along a spectrum, for example, sparse vs. dense dot density or high vs. low tones, with some difficult trials which fall close to the just noticeable difference. Once subjects have learned to make psychophysical discriminations, for which they are rewarded only when correct, they are given an additional response that allows them to avoid taking the test. When subjects opt out of tests, they receive a small, guaranteed reward or an easier subsequent test. If subjects can base responding on metacognitive cues, they should adaptively avoid difficult tests on which they are unlikely to succeed, thus increasing the proportion of trials that lead to reinforcement. Rats, humans, a dolphin, and monkeys selectively avoid difficult primary psychophysical discriminations when given the option (Foote & Crystal, 2007; Smith et al., 1995; Smith, Shields, Schull, & Washburn, 1997).

A persistent question is whether secondary metacognitive responses given in psychophysical tasks could be elicited by public cues imbedded in the primary task rather than by more general private cues such as knowledge state. (Hampton, 2009b; Smith, 2009; Smith, Couchman, & Beran, 2012). Generalization tests, in which the trained metacognitive response is presented with a novel cognitive task, can be used to eliminate the utility of public cues (Smith, Redford, Beran, & Washburn, 2010; Washburn, Smith, & Shields, 2006). When the novel primary task is introduced, many public cues that could have guided apparently metacognitive responding on the previous task are eliminated, including those based on aspects of the stimulus display at test. A subject that bases metacognitive judgments on private cues such as confidence or uncertainty should be able to immediately transfer performance to a new primary discrimination that elicits similar private, cognitive states. In contrast, a subject that exclusively used public cues to make metacognitive judgments on the first task would be unlikely to have immediate success in making judgments with a novel primary task. When the public cues that previously controlled metacognitive responding become unavailable, the subject would lose the basis for his judgments.

Memory tasks provide a stronger evaluation of private cue use because, unlike psychophysical tasks, difficulty is determined by an internal representation of the stimulus seen at study, rather than some aspect of the stimulus display at test (Metcalfe, 2008). Primates have sometimes succeeded at making metamemory judgments (Kornell, Son, & Terrace, 2007; Templer & Hampton, 2012; Washburn, Gulledge, Beran, & Smith, 2010), whereas other nonhumans have not (Brauer, Call, & Tomasello, 2004; Inman & Shettleworth, 1999).

Even when memory tests are used, success on the secondary task does not necessarily rely on attention to private cues. When primary and secondary task responses are presented concurrently, they may be in direct conflict directly with each other. In the tubes task, an information-seeking paradigm, humans or nonhumans are presented with tubes, one of which contains a food reward (Basile, Hampton, Suomi, & Murray, 2009; Call & Carpenter, 2001; Hampton, Zivin, & Murray, 2004). At test, subjects tend to select a response immediately when they have seen baiting, but look inside tubes before their selection when they have not. One alternative explanation to a metacognitive account of this behavior is that nonhumans have a prepotent tendency to reach for food when it is available, a tendency that is not present when they have not seen the food. By this response competition account, subjects are "seeking information" when this and other behaviors are not suppressed by the drive to reach for food. Response competition is not a cue, per se, but could control apparently metacognitive behavior in the absence of private cues.

Other public cues may be taken from a subject's own behavior. For example, a subject that vacillates between multiple choices when it does not have a strong, immediate response could learn that this behavior predicts unsuccessful test responses. Similarly, nonhumans could learn to associate hesitation or long latency to take the test with poor performance. Although these cues are self-generated, they are nonetheless publically observable, and do not demand private access to the subject's internal cognitive state. Public cues that relate to the subject's behavior at test can be eliminated by structuring the task such that the metacognitive judgment is presented before test stimuli are available (Hampton, 2001). Psychophysical task difficulty relies on the stimuli available at test, so prospective judgments are only available for memory tests, which could elicit a private discriminative cue, such as memory strength.

It is challenging to resolve which mechanisms underlie nonhumans' apparently metacognitive performance because different paradigms are often used in isolation. If monkeys succeed at a task that seems to necessitate reliance on a private cue, e.g., prospective metamemory judgment, we may assume that they possess the capacity to attend to private cues. However, it remains unclear whether they also attend to a private cue in service of a task that contains more obvious public cues, such as a psychophysical discrimination. Further, the existing work cannot demonstrate whether the private cue used to solve a prospective memory judgment is task-specific, as in memory strength, or reflects an assessment of knowledge state that would generalize to other tasks. Similar performances may be achieved through different mechanisms. Some tasks can be solved by multiple means, and it is possible that animals use different cues under different circumstances. Public cues are overt and may be more salient than comparatively subtle assessments of private knowledge states. However, it seems improbable that subjects would spontaneously and rapidly shift cues under novel circumstances. If monkeys are able to rapidly generalize metacognitive responding across multiple perceptual tasks, it narrows the range of information available in making a metacognitive discrimination to cues shared across tasks. If monkeys further generalize to memory tasks, which are more likely to demand attention to private cues, it would suggest reliance on private cues across tasks.

In the current experiments, we evaluated bases of metacognitive responding in monkeys by implementing a variety of transfer tasks that differed in the availability of public cues that could be used to predict task performance. We assessed immediate generalization of the metacognitive response. In Experiment 1, we trained a group of monkeys on a perceptual discrimination and then provided a secondary metacognitive "*decline-test*" response which allowed them to selectively avoid certain trials for a guaranteed small reward. In Experiments 2-4, we used 3 novel perceptual tasks to evaluate whether monkeys could generalize use of the *decline-test* response in the absence of stimulus-specific cues from their initial training. In Experiment 5, we evaluated the possibility that monkeys based their metacognitive judgments on public cues that were general across the perceptual tasks by training them on a delayed matching-to-sample task, which is based on memory and therefore not bound to external stimuli. In Experiment 6, we evaluated the monkeys' ability to make metacognitive judgments without the availability of any public cues associated with the test by presenting the metacognitive judgments prospectively, before the test. If metacognitive responding is under the control of private cues, such as assessment of knowledge states, we expect that monkeys will show immediate generalization of the *decline-test* response across perceptual, concurrent choice, and prospective choice memory tasks. If monkeys base their judgments on public cues, such as their own behavior or some aspect of the stimulus display at test, we expect that they will fail to generalize adaptive use of the *decline-test* response to novel perceptual and memory tasks.

2. Subjects

Subjects were 12 pair-housed male rhesus macaque monkeys (*Macaca mulatta*), average age 5.6 years at the beginning of these studies, with a one year history of computerized cognitive testing. Six subjects had previous experience with a manual metacognition task (Templer & Hampton, 2012).

3. Apparatus

We tested monkeys in their home cages, using portable touch-screen computer rigs consisting of a laptop computer (Dell, Round Rock, TX) with generic speakers, a 15" color LCD touchscreen (ELO, Menlo Park, CA), and two automated food dispensers (Med Associates Inc., St. Albans, VT) that dispensed into food cups beneath the screen. Food reinforcement consisted of 94 or 97 mg nutritionally complete primate pellets (Bio-Serv, Frenchtown, NJ and Purina TestDiet, Richmond, IN). We presented stimuli and collected responses using programs written in Presentation (Neurobehavioral Systems, Albany, CA).

4. Procedure

4.1 Monkey housing and testing conditions

During testing, paired monkeys were separated by dividers that allowed visual and physical contact through large slots, but prevented access to adjacent testing equipment. Monkeys had access to their testing rigs up to seven hours per day, 6 days per week. Each day, monkeys participated in 2-4 consecutive experiments, one of which was the study reported here. Eight monkeys received a full food ration daily. The other four monkeys were on caloric restriction for part of this study. All dietary changes for these monkeys were supervised by veterinary staff and weights were monitored weekly. Water was available ad libitum.

4.2 Training on perceptual discriminations

We trained monkeys on a series of perceptual discrimination tasks. Within each task, the target stimulus remained the same across trials, and difficulty was varied on a trial by trial basis by changing the discriminability of three identical distracters from the target. To start a trial, monkeys touched a green ready square at the bottom center of the screen. All responses required two touches (FR2) to prevent recording undirected contacts. The target and the three identical distracters then appeared in the four corners of the screen. Within each task, the distracters differed from the target by 5 levels of difficulty along one stimulus dimension (size, brightness, arc length, or degrees rotation). Each difficulty level consisted of two different distracter values, one lesser (e.g. dimmer) and one greater (e.g. brighter) in magnitude than the target by equal amounts. In the final phase of each experiment, distracters identical to the target were used on the hardest trials. On such impossible trials, one location was still selected pseudo-randomly as the "target", and selection of this location resulted in reward as on correct trials.
Choice of the target resulted in a distinctive auditory signal and food reinforcement. Selection of a distracter resulted in auditory feedback and black screen for a timeout period. A 2-second inter-trial interval separated consecutive trials.

Each session consisted of 100 trials, with 20 trials from each difficulty level, half from each distracter that represented that difficulty level. Difficulty level and target location were pseudo-randomly intermixed within a session to maintain counterbalancing.

5. Experiment 1- Size Discrimination

5.1 Training on size discrimination

For the first task, monkeys were required to select a target from distracters on the basis of size. Stimuli were otherwise identical circles that differed in size. The target circle was a constant size. Over three phases of increasing difficulty (Table 1), monkeys were trained to discriminate the target from distracters (Figure 1).

Figure 1. Stimuli used in the final phase of the size discrimination (top). Labels on the circles indicate actual diameters, in pixels, used in the experiment, but were not shown to the monkeys. At each level of difficulty (bottom), there were two absolute distracter sizes. The easiest distracters (i.e., difficulty level 1) were 24 pixels larger or smaller in diameter than the target, and the hardest (i.e., difficulty level 5) identical to the target.

To effectively titrate the difficulty of the task, we carried out training in three phases, each with different distracter sizes. Subjects worked on each phase until they had completed at least five 100-trial sessions, with 85% accuracy on the easiest level of distracters for two consecutive sessions. After completion of the third phase of training on the size discrimination, monkeys began training on use of the *decline-test* response.

Table 1. Distracter size, diameters in pixels

	easy				hard	target	hard				easy
Phase1	50	54	58	62	66	100	134	138	142	146	150
Phase 2	60	68	76	84	92	100	108	116	124	132	140

Phase 3	76	82	88	94	100	100	100	106	112	118	124

5.2 Training on the decline-test response

Trials proceeded as described above, except that two additional black and white clipart choice stimuli could be displayed concurrently with test stimuli. The *accept-test* stimulus, a check-marked square, occupied the right center of the screen. Touches to the *accept-test* stimulus extinguished choice stimuli and activated test stimuli on screen. Choice of the target resulted in a distinctive auditory signal and two food pellets. Selection of a distracter resulted in auditory feedback and black screen for a timeout period. The *decline-test* stimulus, a thumbs-down, occupied the left center of the screen. Selection of the *decline-test* stimulus resulted in the immediate presentation of a red rectangle at the top center of the screen. Touches to this *guaranteed small reward* stimulus resulted in a distinctive auditory signal and one food pellet. Initially, all timeouts were 500-miliseconds. As training progressed, some monkeys that failed to make appropriate use of the *decline-test* response were given longer timeouts to increase attention to the difficulty of different trial types. Timeouts eventually ranged from 500-miliseconds to 240-seconds, according to individual learning and motivation to use the *decline-test* response.

On 2/3 of trials, subjects were presented with both choice stimuli (Figure 2, left side). On the other 1/3 of trials, only the *accept-test* stimulus was presented, forcing subjects to take the test (Figure 2, right side). Chosen and forced trials were evenly distributed across difficulty conditions. Subjects were trained until they had completed at least 20 sessions and showed at least 30% difference in use of the *decline-test* response on easiest and hardest trials averaged across 5 sessions. Each session contained 180 trials, with 36 trials from each difficulty level, half larger and half smaller than the target. Difficulty level and target location were pseudo-randomly intermixed within a session.

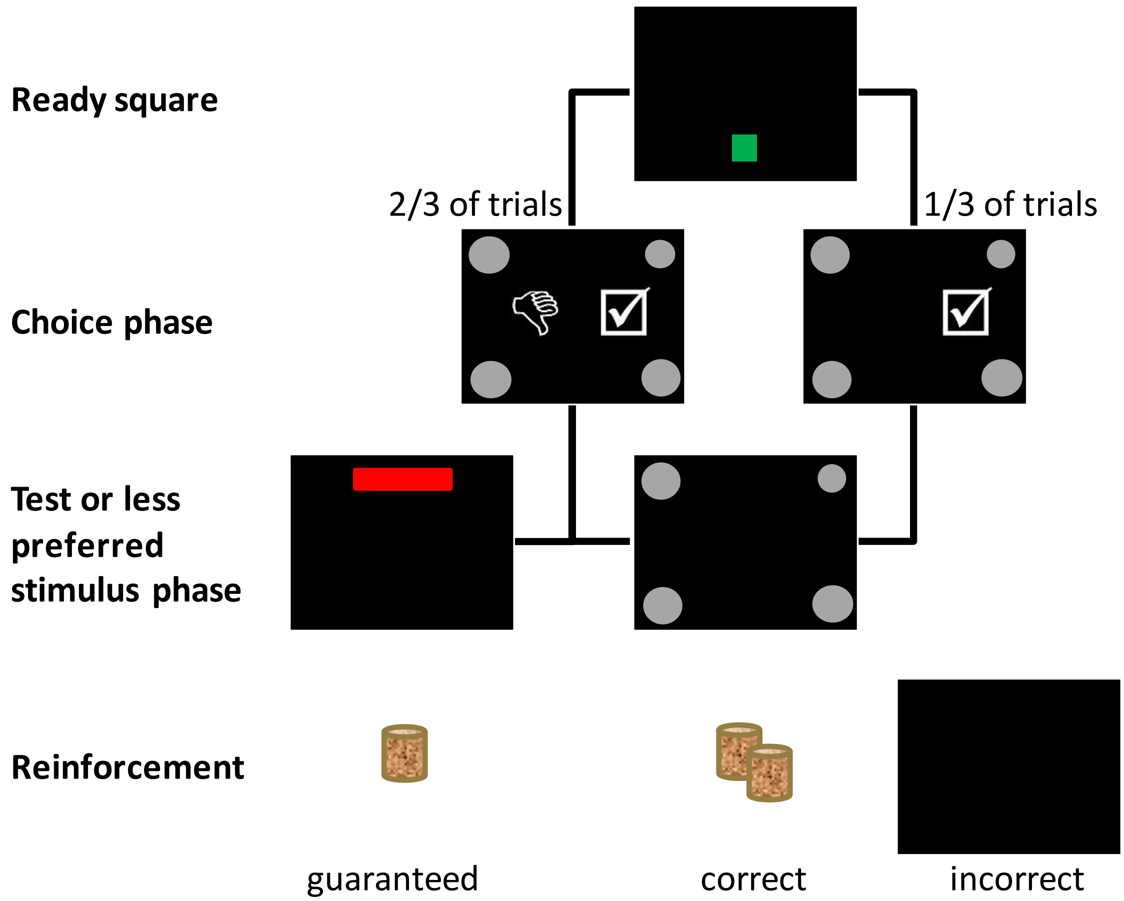


Figure 2. Steps to complete a trial of the training task with choice stimuli present. Monkeys touched the green ready square to initiate trials. Choice and test images then appeared on screen: the target circle of constant size, three distracters identical to one another, and the *accept-test* and *decline-test* choice stimuli. On 1/3 of trials (right), the *decline-test* response did not appear. Choice of the *accept-test* stimulus extinguished choice stimuli and activated test stimuli. Tests resulted in food reinforcement of two pellets (correct) or a black time out screen (incorrect). Selection of the *decline-test*

response caused the *guaranteed small reward* stimulus screen to appear. Touches to this stimulus resulted in guaranteed food reinforcement of one pellet.

5.3 Data analysis

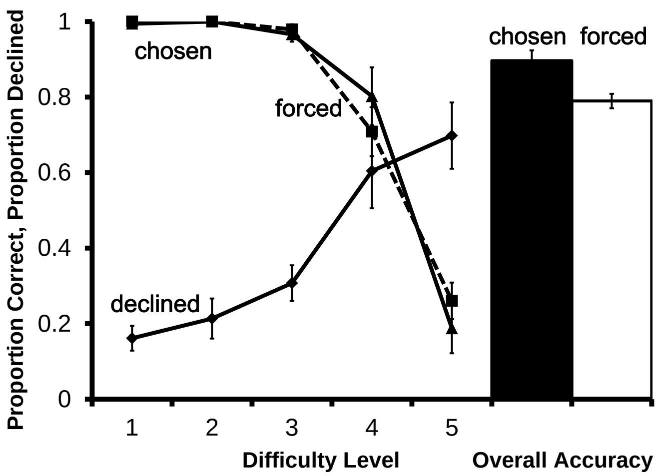
All proportions were arcsine transformed before statistical analysis to better approximate the normality assumption underlying parametric statistics (Keppel and Wickens 2004, p. 155). Geisser–Greenhouse correction was used, and appropriately adjusted degrees of freedom reported, whenever the sphericity assumption was violated (Keppel and Wickens 2004, p. 378).

For all experiments, we assessed differences in accuracy between forced and chosen tests, pooled across all levels of difficulty, using paired t-tests. We assessed accuracy and use of the *decline-test* response as functions of difficulty level using repeated measures ANOVA. Follow-up planned paired t-tests were used to compare accuracy and use of the *decline-test* response between difficulty levels 1 and 5.

5.4 Results and discussion

Eight of twelve subjects reached criterion with the *decline-test* response. The following analysis is based on their performance on the final criterion session. The other four monkeys never learned to use the *decline-test* response adaptively for this task. The four monkeys that were unable to reach criterion in Experiment 1 were dropped from all subsequent analyses.

Monkeys' accuracy differed as a function of the similarity between the size of the distracters and the target (Figure 3; $F_{4,28}$=61.94, $p<.001$) . Monkeys were significantly more accurate on difficulty level 1 trials than on difficulty level 5 trials (T_7 = 13.41, p<.01).



.

Figure 3. Performance of 8 monkeys on the final criterion session of the size discrimination with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

We were able to train eight monkeys to criterion based on use of the *decline-test* response. Monkeys' use of the *decline-test* response differed as a function of task difficulty ($F_{1.415,9.90}$=17.80, p<.01). Monkeys declined significantly more trials from difficulty level 5 than level 1 (T_7 =-6.25, p<.01).

Each session, subjects were required to take the same number of forced tests across difficulty levels; however, the proportion of chosen tests from each difficulty level varied according to subjects' use of the *decline-test* response. Because subjects selectively avoided more difficult trials, overall chosen test accuracy disproportionately reflects performance on difficulty level 1 trials compared with difficulty level 5 trials. Whereas overall forced test accuracy always represents 12 trials at each difficulty level, overall chosen accuracy for a single subject could, for example, consist of 24 trials at difficulty level 1 and 0 trials from difficulty level 5. Thus, monkeys increased the proportion of trials resulting in reinforcement by declining the most difficult trials when given the option. Overall accuracy on chosen tests was significantly higher than on forced tests ($T_7 = 3.23$, $p < .05$).

Forced and chosen accuracy functions did not differ from one another ($F_{1,7} = .08$, $p = .79$). Because difficulty is dictated by a difference in perceptibility, it seems that there would be little variation between trials on which monkeys do or do not succeed. Additionally, if subjective difficulty is determined in the context of the session, the selected difficulty levels likely provide a far more overt cue than any subtle variation within difficulty level.

Eight monkeys made selective use of the *decline-test* response consistent with metacognition in Experiment 1. Because monkeys received extensive training with the choice stimuli in the context of this size discrimination test, they may have used some specific aspect of the display to make their decision. For example, the distance between stimuli and the edge of the screen, the overall luminance of the display, or some other feature could have provided a cue as to the likelihood of reinforcement. To address this potential concern, we conducted a transfer test in Experiment 2.

6. Experiment 2- Brightness Discrimination

6.1 Rationale

Monkeys received extensive training with the *decline-test* response in the context of the size discrimination used in Experiment 1. Due to this training, we cannot conclusively say what type of cue they monitored to generate the observed difference between *decline-test* use on difficulty levels 1 vs. 5. It is possible that monkeys made apparently metacognitive choices on the basis of learned associations between screen displays and probability of reinforcement. Generalization tests provide a means to assess the type of cues that animals use to make secondary metacognitive judgments because changing the primary task eliminates public cues specific to the original task. If monkeys used private cues to guide their use of the *decline-test* response in Experiment 1, they should be able to immediately transfer this performance to a novel brightness discrimination in Experiment 2, which we expect to elicit similar private, cognitive states. In contrast, if public cues exclusively controlled the pattern of performance in Experiment 1, the monkeys will be unlikely to succeed at the rapid generalization of the metacognitive response to a new primary discrimination in Experiment 2.

6.2 Subjects

The eight monkeys who met criterion with the *decline-test* response in Experiment 1 were used in Experiment 2.

6.3 Training on brightness discrimination

Monkeys were required to select a target from distracters on the basis of brightness. Stimuli consisted of greyscale squares that differed in brightness, but were identical along all other dimensions, with difficulty varied according to the same scheme used in Experiment 1 (Figure 4; Table 2).

Figure 4. Stimuli used in the final phase of the brightness discrimination (top). Labels on the squares indicate actual RGB values used in the experiment. The difficulty level 1 stimuli were 64 RGB brighter or darker than the target.

Table 2. Distracter brightness in RGB values

	easy				hard	target	hard				easy
Phase 1	64	74	84	94	104	128	152	162	172	182	192
Phase 2	64	74	84	94	128	128	128	162	172	182	192

6.4 Pre-transfer review

Prior to transfer, we assessed monkey performance on the size discrimination task from Experiment 1 and performance on the new brightness discrimination. Monkeys had to complete a 180-trial session of the size discrimination. Then, monkeys had to complete a session of the brightness discrimination without the *decline-test* response available. Monkeys had to complete this cycle at least five times (10 sessions total). In order to proceed to the transfer task, they had to demonstrate in consecutive sessions a 30% difference in the use of the *decline-test* response between difficulty levels 1 and 5 for size discriminations and 85% accuracy on level 1 trials of the brightness discrimination across the last 2 sessions. This ensured that the earlier pattern of *decline-test* responding was intact and that the brightnesss discrimination had sufficient variation in difficulty to elicit both *decine-test* and *accept-test* responses.

6.5 Transfer of the decline-test response

Trial contingencies were the same as those described for Experiment 1.

6.6 Data analysis

To assess rapid generalization of metacognitive responding, we analyzed only the first session of the new discrimination for which the *decline-test* response was available.

6.7 Results and discussion

Monkeys' accuracy differed as a function of the similarity between the color of the distracters and the target (Figure 5; $F_{4,28}$=52.01, p<.01); monkeys were significantly more accurate on the difficulty level 1 trials than on difficulty level 5 trials (T_7 =-12.58, p<.01).

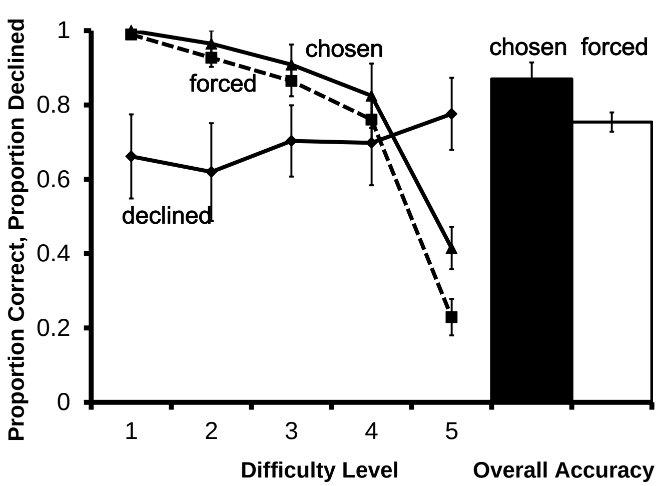


Figure 5. Performance of 8 monkeys on the first session of the brightness discrimination with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

In the first transfer session, use of the *decline-test* response showed a slight trend toward a relation with task difficulty, but the effect was not statistically significant ($F_{1.65,11.53}$=3.26, p=.08). Monkeys did not use the *decline-test* response more at difficulty level 5 trials compared with difficulty level 1 trials (T_7 =-1.72, p=.13). Overall use of the *decline-test* response was high in Experiment 2 compared to Experiment 1 (69.17% of trials on which the response was available in Experiment 2, compared to a mean of 39.69% in Experiment 1). Two monkeys declined all trials for which they had the option. When these animals are excluded from the analysis, the remainder of the group shows a significant relation between difficulty level and use of the *decline-test* response ($F_{4,20}$=3.67, p<.05). High use of the *decline-test* response may have been a reaction to task novelty, as this response had only previously been available with one type of discrimination. Another possibility is that this pattern of responding reflected a discrepancy in perceived difficulty between the two tasks.

The difference between overall forced and chosen accuracy was not statistically significant (T_5 =-2.13, p=.09). The two monkeys that declined all trials for which they had the option did not have data for chosen accuracy available for comparison.

As a group, the monkeys did not successfully generalize adaptive use of the *decline-test* response to the novel perceptual discrimination. However, they did show a trend toward a relation between difficulty level and use of the *decline-test* response, and some individual monkeys appear to have generalized successfully. Task novelty combined with the high overall use of the response may indicate that monkeys need exposure to the *decline-test* response in multiple contexts to successfully transfer it to novel tasks. Response generalization may improve after multiple exposures in diverse contexts. To provide further practice with the *decline-test* response under multiple contexts, we extended training and provided another novel generalization task, described in the next experiment.

7. Experiment 3- Arc Length Discrimination

7.1 Rationale

The generalization test in Experiment 2 provided ambiguous results. Several monkeys showed robust transfer to the new task, and 5 of 8 monkeys declined numerically more difficulty level 5 than difficulty level 1 tests. However, the effect of difficulty level on use of *decline-test* response was non-significant for the group. Overall use of the *decline-test* response was very high, at least in part because 2 monkeys declined all tests for which the option was available, a result which is uninterpretable. Given these results, and the possibility that additional generalization opportunities may reduce any surprise toward seeing metacognitive choices in new contexts, we provided another transfer task, an arc length discrimination, in Experiment 3.

6.2 Subjects

Seven of the eight monkeys from Experiment 2 participated in Experiment 3. One monkey was removed from the laboratory for failure to complete trials in a reasonable amount of time. He had successfully met criterion for use of the *decline-test* response in Experiment 1 and used the *decline-test* response for all available trials in Experiment 2. He is not included in this or any subsequent analyses.

6.3 Pre-training on known discriminations

Following the initial transfer session in Experiment 2, monkeys were required to cycle through the size and brightness discriminations with the secondary metacognitive task, in the order described in the pre-transfer training for that experiment, until they demonstrated a 30% difference between difficulty levels 1 and 5 in use of the *decline-test* response on the brightness discrimination. This ensured that the pattern of *decline-test* responding from Experiment 1 was intact, a necessary foundation for a subsequent transfer task.

7.4 Training on arc length discrimination

Monkeys were required to select a target from distracters on the basis of length. Stimuli consisted of arcs that differed in length, but were identical along all other dimensions (Figure 6; Table 3). Training was otherwise the same as in Experiments 1 and 2. One monkey's performance was still at chance after 20 sessions of training. At this point, he was given distracters more discriminable from the target, to make the task easier (Table 3 B).

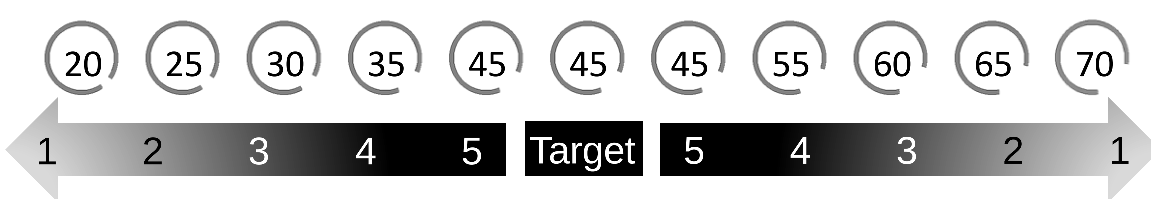


Figure 6. Stimuli used in the final phase of the arc length discrimination (top). Labels on the arcs indicate actual length of distracters, given in degrees missing from the circle, that

were used in the experiment. The difficulty level 1 stimuli were 25 degrees longer or shorter than the target.

Table 3. Distracter size, degrees of gap missing from circle

	easy				hard	target	hard				easy
Phase 1	70	65	60	55	50	45	40	35	30	25	20
Phase 2	70	65	60	55	45	45	45	35	30	25	20

Table 3 B.

	easy				hard	target	hard				easy
Phase 1	80	75	70	65	60	45	30	25	20	15	10
Phase 2	80	75	70	65	45	45	45	25	20	15	10

7.5 Pre-transfer review

We assessed monkey performance on the size, brightness, and arc-length discriminations. Monkeys had to complete a 120-trial session of the size discrimination followed by a 120-trial session of the brightness discrimination, both with *decline-test* response available, as described in Experiments 1 and 2. Then, monkeys had to complete a session of the arc length discrimination. Monkeys had to complete this cycle at least eight times (24 sessions total). In order to proceed to the transfer task, they had to demonstrate in consecutive sessions a 30% difference in the use of the *decline-test* response between difficulty levels 1 and 5 for size and brightness discriminations and 85% accuracy on level 1 trials of the arc length discrimination across the last 2 sessions. This ensured that the earlier pattern of *decline-test* responding was intact and that the arc-length discrimination had sufficient variation in difficulty to elicit both *decine-test* and *accept-test* responses.

7.6 Transfer of the decline-test response

Trial contingencies were the same as those described for Experiment 1.

7.7 Results and discussion

Monkeys' accuracy differed as a function of the similarity between the length of the distracters and the target (Figure 7; $F_{4,24}$=29.47, p<.01); monkeys performed significantly more accurately on the easiest trials than impossible ones (T_6 =9.16, p<.01).

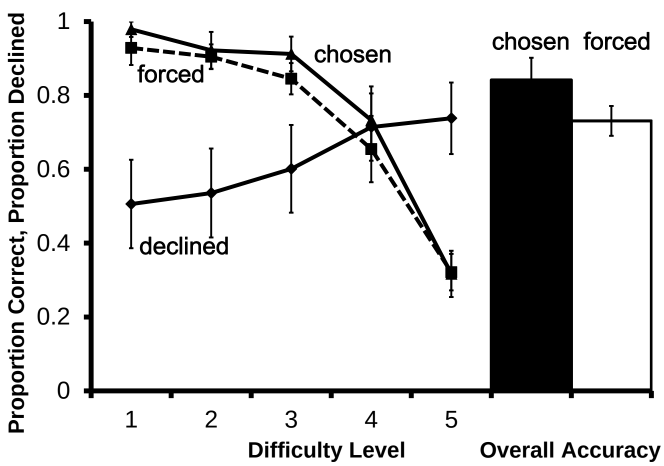


Figure 7. Performance of 7 monkeys on the first session of the arc length discrimination with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

Use of the *decline-test* response showed a trend toward a relation with task difficulty, but the effect was not statistically significant ($F_{1.23,7.40}$=4.59, p=.06); however, monkeys used the *decline-test* response significantly more frequently on difficulty level 5 than difficulty level 1 trials (T_6 =2.50, p<.05).

Overall performance on chosen tests was higher than on forced, but this difference was not significant (T_5 =1.40, p=.22). One monkey declined all trials for which this option was available, and therefore did not have chosen accuracy available for comparison.

8. Experiment 4- Rotation Discrimination

8.1 Rationale

Experiment 3 provided stronger, but still not robust, evidence of transfer. Use of the *decline-test* response as a function of difficulty level approached significance, and monkeys declined significantly more difficulty level 5 than difficulty level 1 trials. Though monkeys showed improved evidence of transfer, they still did not show the robust transfer that would be associated with use of private metacognitive cues. Given the fact that the monkeys appeared to be increasing their proficiency with the *decline-test* response, we provided an additional opportunity for transfer in Experiment 4, with a rotation discrimination.

8.2 Subjects

The seven monkeys from Experiment 3 participated in Experiment 4.

8.3 Pre-training on known discriminations

Following the initial transfer session in Experiment 3, monkeys were required to cycle through size, brightness, and arc length discriminations with the secondary metacognitive task, in the order described in the pre-transfer training for that experiment, until they demonstrated a 30% difference between difficulty levels 1 and 5 in use of the *decline-test* response on the arc-length discrimination. This ensured that the pattern of *decline-test* responding from Experiment 1 was intact, a necessary foundation for a subsequent transfer task.

8.4 Training on rotation discrimination

In Experiment 4, monkeys were required to select a target from distracters based on degrees of rotation. Stimuli consisted of circles containing an array of small dots. Stimuli differed in rotation from the center point of the outer circle, but were identical along all other dimensions. The target stimulus was held at a constant rotation (Figure 8; Table 4).

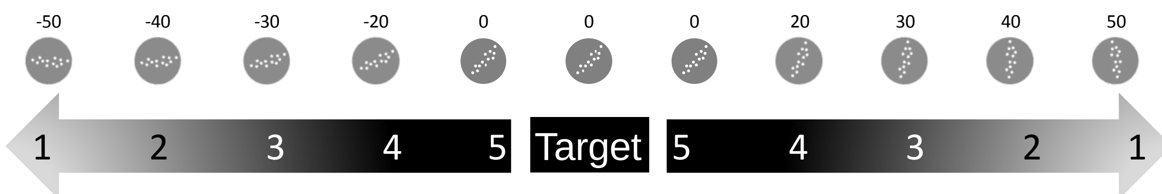


Figure 8. Stimuli used in the final phase of the rotation discrimination (top). Labels on the stimuli indicate actual rotation, in degrees, used in the experiment. The difficulty level 1 stimuli were rotated 50 degrees from the target, either clockwise or counter-clockwise.

The initial dot array used made it very difficult for monkeys to discriminate the target from the rotated distracters. The first two monkeys on the task continued with these stimuli. Subsequent subjects received stimuli with a greater number of dots, organized in a more linear pattern, as shown here.

Table 4. Distracter rotation from target, in degrees

	easy				hard	target	hard				easy
Phase 1	-50	-40	-30	-20	-10	0	10	20	30	40	50
Phase 2	-50	-40	-30	-20	0	0	0	20	30	40	50

8.5 Pre-transfer review

We gave monkeys at least one 180-trial review session of each of the size, brightness, and arc-length discriminations with the *decline-test* response available as described in Experiment 1 to ensure that they maintained their prior appropriate use of the *decline-test* response. Over these trials, monkeys were required to show a 30% difference between *decline-test* response use on difficulty levels 1 and 5 on at least one prior task. Review sessions alternated with sessions of the rotation discrimination, for which monkeys were required to maintain 85% accuracy on difficulty level 1 trials. This ensured that the earlier pattern of *decline-test* responding was intact and that the rotation discrimination had sufficient variation in difficulty to elicit both *decine-test* and *accept-test* responses.

8.6 Transfer of the decline-test response

Trial contingencies were the same as those described for Experiment 1. Ceiling use of the *decline-test* response, exhibited by some monkeys in prior experiments, could obscure an effect of difficulty level. To prevent ceiling use of the *decline-test* response, we retitrated the number of required touches to the *guaranteed small reward stimulus* for the monkey who declined all tests in Experiment 3. Following a session when he declined over 70% of trials, the number of required touches to the stimulus to obtain the guaranteed food reward was doubled. Following a session when he declined fewer than 30% of trials, this number was halved.

8.7 Results and discussion

Monkeys' accuracy differed as a function of the similarity between the degrees rotation of the distracters and the target (Figure 9; $F_{4,24}$=23.99, p<.01); monkeys performed significantly more accurately on difficulty level 1 than level 5 (T_6 = 7.50, p<.01).

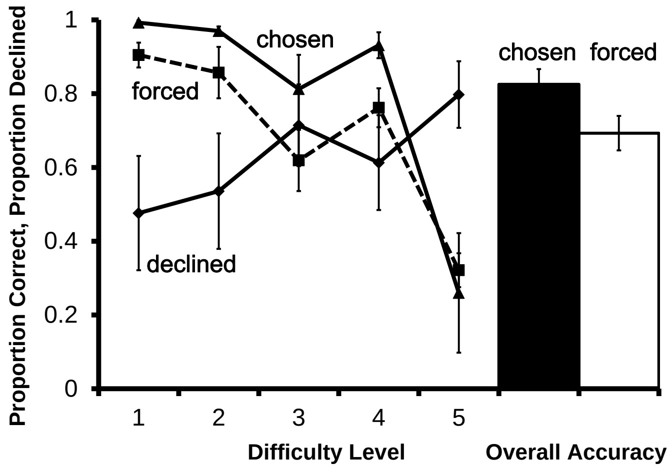


Figure 9. Performance of 7 monkeys on the first session of the rotation discrimination with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

Use of the *decline-test* response differed as a function of difficulty level ($F_{1.432,8.592}$ =5.93, p<.05), and monkeys declined significantly more difficulty level 5 trials than difficulty level 1 (T_6 = 2.81, p<.05). This was the first task on which monkeys showed robust, successful generalization of the *decline-test* response, indicated by its use as a function of trial difficulty level.

Monkeys increased the proportion of trials resulting in reinforcement by declining the most difficult trials when given the option. Overall performance on chosen tests was significantly higher than performance on forced tests (T_5 =-3.07, p<.05), indicating that differential decline of difficult trials improved overall performance. One monkey declined all trials for which this option was available, and therefore did not have chosen accuracy available for comparison.

Transfer of the adaptive use of the *decline-test* response was most robust for this final perceptual task. There are several possible reasons for the monkeys' increased ability to generalize, although this list is not exhaustive. First, transfer may have improved as monkeys completed multiple tasks simply because they increased expertise with the *decline-test* response as they got more practice with it, in terms of both time and number of trials. Because of subtle differences in training and subjective differences in perceived task difficulty, we cannot assess this possibility empirically. Another possibility is that the meaning of the *decline-test* response shifted and evolved as monkeys were forced to apply it to multiple perceptual domains. A final potential explanation is that the surprise of seeing the *decline-test* response in a new context was initially jarring enough to interfere with performance. This could lead to its excessive use, which we saw especially in the first transfer task. After repeated exposures in multiple novel contexts, this surprise would gradually lessen, and performance on transfer tests would concurrently improve. Regardless of its cause, use of the *decline-test* response on this final perceptual task shows immediate generalization, suggesting that monkeys used a cue that was available across perceptual domains. Monkeys could have used a cue related to reinforcement probabilities or some general aspect of stimulus display, or they could have used a private cue. To further evaluate the type of cue monkeys use to make the *decline-test* response the next transfer task involved memory.

9. Experiment 5- Concurrent Metamemory

9.1 Rationale

Following robust transfer performance in the final perceptual task, the rotation discrimination described in Experiment 4, we wanted to evaluate whether this apparently metacognitive responding relied on some common aspect of all of the prior perceptual tasks. Because Experiment 5 uses a memory task, use of public cues available for prior tasks, such as direct comparison of visual stimuli, would not be an effective strategy. Our memory tasks use the same four images on every test, so the appearance of the stimuli should not provide any useful cues about task difficulty. If earlier performance reflects reliance on a private cue, based on the monkey's cognitive state, we expect them to successfully transfer use of the *decline-test* response to the memory domain. If monkeys used a public cue that was available across the comparatively similar perceptual discriminations, they will not have the basis to successfully transfer use of the *decline-test* response to this novel testing scenario.

9.2 Memory tasks

We trained monkeys on a memory task to assess generalization of the *decline-test* response to a non-perceptual domain. We used a delayed matching-to-sample task (DMTS). A small, familiar set of 4 clipart images was used across all sessions, such that every image was seen on every trial. To start a trial, monkeys touched a green ready square at the bottom center of the screen. A sample image then appeared in one of the four corners of the screen. Touches to this image resulted in a blank screen for a delay. Choice of the sample image seen prior to the delay resulted in a distinctive auditory signal and food reinforcement. Selection of a distracter resulted in auditory feedback and black screen for a timeout period. A 5-second inter-trial interval separated consecutive trials. Sample image, target location, and delay length were balanced and pseudo-randomized within each session. During training, 6,12 ,24 ,and 48- second delays were intermixed in each session.

Following training in the matching-to-sample procedure, monkeys were given the opportunity to transfer use of the *decline-test* stimulus to the memory test. Contingencies were the same as described for perceptual discriminations. On memory tasks, the choice stimuli could appear concurrently with the test as described in Experiment 5, or prospectively before the test as described in Experiment 6 (Figure 10).

Figure 10. Steps to complete a trial of the memory task with metacognitive choice stimuli. Monkeys touched the green ready square to initiate trials (not shown). A sample clipart image then appeared on screen. On 2/3 of trials, the *decline-test* and *accept-test* stimuli appeared after a delay. Because difficulty on a specific test relies on memory strength through a delay, rather than perceptual comparison at test, memory tasks allow us to manipulate the amount of information available at the time of the metacognitive

choice. The choice stimuli could appear concurrently with the test (above) or prospectively before the test (below).

9.3 Subjects

The seven monkeys from Experiment 4 participated in Experiment 5.

9.4 Pre-training on known discriminations

Following the initial transfer session in Experiment 4, monkeys were required to cycle through size, brightness, arc-length, and rotation discriminations with the secondary metacognitive task, in the order described in the pre-transfer training for that experiment, until they demonstrated a 30% difference between difficulty levels 1 and 5 in use of the *decline-test* response on the rotation discrimination. This ensured that the pattern of *decline-test* responding from Experiment 1 was intact, a necessary foundation for a subsequent transfer task.

9.5 Training for Memory Task

Prior to Experiment 5, monkeys completed 20 sessions of DMTS. After these training sessions, an additional delay length was added, so that delays lasted .2, 6, 12, 24, or 48-seconds. Difficulty level for this task was based on delay length, such that difficulty level 1 trials included a .2-s delay, difficulty level 2 trials included a 6-s delay, etc.

9.6 Pre-transfer review

We gave monkeys at least one review session of each of the prior perceptual discriminations with the *decline-test* response available as described in Experiment 1. Over these trials, monkeys were required to show a 30% difference between *decline-test* response use on difficulty levels 1 and 5 on at least one prior task to ensure that they maintained their prior appropriate use of the *decline-test* response. Review sessions alternated with sessions of the DMTS, for which monkeys were required to maintain 85% accuracy on difficulty level 1 trials. Monkeys that demonstrated accuracy below 85% on difficulty level 1 trials after 4 sessions were given 10 remedial sessions of DMTS only. If accuracy was still below 85% at the end of this remedial block, ITI was increased by 5-seconds to decrease interference from other trials. Performance was re-evaluated every 4 sessions, at which time ITI was increased by 5-second intervals or monkeys were returned to pre-transfer review.

9.7 Transfer of the decline-test response

Trial contingencies were the same as those described for Experiment 1. The monkey who declined all of the trials in Experiment 4 was given the changing FR to obtain his small guaranteed reward, as described in Experiment 4. The monkey who previously experienced this contingency did so again in Experiment 5. As in prior experiments, in Experiment 5, the choice stimuli were presented concurrently, at the same time as test stimuli.

9.8 Results and discussion

Monkeys' accuracy differed as a function of the delay length (Figure 11; $F_{4,24}=3.12$, $p<.05$); monkeys performed significantly more accurately on difficulty level 1 than level 5 ($T_6 = 2.45$, $p=.05$).

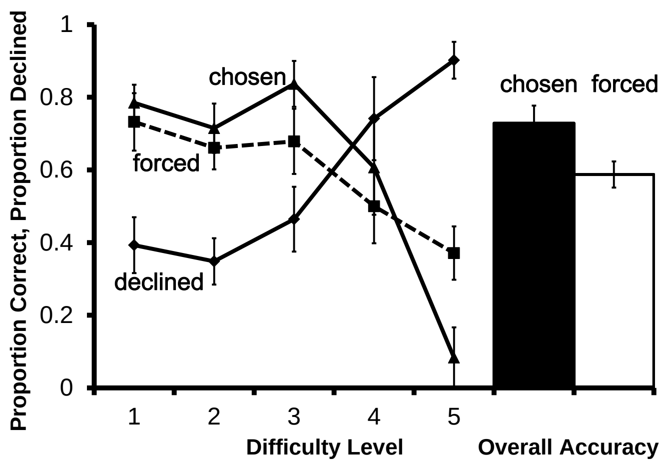


Figure 11. Performance of 7 monkeys on the first session of the concurrent metamemory task with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

Use of the *decline-test* response differed as a function of task difficulty ($F_{4,24} = 23.94$, $p<.01$), and monkeys declined significantly more difficulty level 5 trials than difficulty level 1 ($T_6 = -7.77$, $p<.01$).

Overall performance on chosen tests was higher than performance on forced tests, although the difference was not statistically significant ($T_6 = -2.29$, $p=.06$).

Monkeys successfully transferred adaptive use of the *decline-test* response to the novel context of a memory test. This immediate generalization indicates that the use of the *decline-test* response did not depend on public characteristics shared across the perceptual discriminations.

10. Experiment 6- Prospective Metamemory

10.1 Rationale

Even when memory tests are used, the metacognitive judgment could be completed without the use of private cues. Public cues in response to the items presented at test, such as vacillation between multiple responses, could be consistent across multiple perceptual and memory domains. For example, if a monkey always selects the *decline-test* response when he is alternating between multiple potential answers, he could make judgments based on the public cue of his own behavior across Experiments 1-5. Memory tests allow presentation of the metacognitive choice in the absence of the test, eliminating cues associated with the test. For Experiment 6, the metacognitive choice was presented prospectively, before subjects saw the test. By forcing monkeys to choose whether to take the test before they saw it, we further limited the availability of public discriminative cues. Prospective judgments eliminate responses and behaviors bound to the appearance of the test, and are therefore useful in demonstrating whether use of the *decline-test* response can be guided by private cognitive cues.

10.2 Subjects

We assessed transfer for six of the seven monkeys from Experiment 5. One monkey included in prior experiments was removed from this study because he did not maintain above-chance forced memory test performance.

10.3 Pre-training on known discriminations

Following the initial transfer session in Experiment 5, monkeys were required to cycle through all known discriminations with the secondary metacognitive task, in the order described in the pre-transfer training for that experiment, until they demonstrated a 30% difference between difficulty levels 1 and 5 in use of the *decline-test* response on the concurrent memory task. This ensured that the pattern of *decline-test* responding from Experiment 1 was intact, a necessary foundation for a subsequent transfer task.

10.4 Pre-transfer review

Prior to the transfer task, monkeys were required to complete at least 5 sessions of prospective choice DMTS with all forced trials. These sessions were intended to familiarize monkeys with completing the choice phase before seeing the test, a change which could have been distracting or confusing if first seen at transfer. Monkeys were required to maintain 85% accuracy on difficulty level 1 trials to proceed.

We gave monkeys at least one review session of the rotation discrimination with the *decline-test* response available as described in Experiment 1. Over these trials, monkeys were required to show a 30% difference between *decline-test* response use on difficulty levels 1 and 5, to ensure the maintenance of prior appropriate use of the *decline-test* response.

10.5 Transfer of the decline-test response

In Experiment 6, the choice stimuli were presented prospectively before test stimuli. Trial contingencies were the same as those described for Experiment 1. Both monkeys who experienced a changing FR to receive guaranteed reward continued to experience this contingency.

10.6 Results and discussion

Monkeys' accuracy differed as a function of the delay length (Figure 12; $F_{4,20}$=6.70, p<.01); monkeys performed significantly more accurately on difficulty level 1 than level 5 (T_5 = 4.37, p<.01).

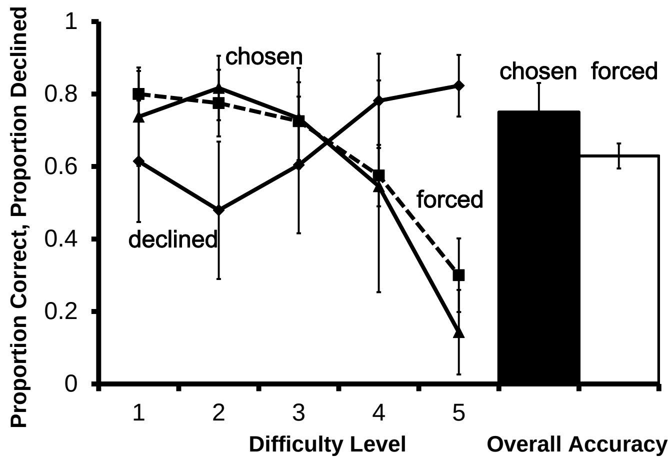


Figure 12. Performance of 6 monkeys on the first session of the prospective metamemory task with the *decline-test* response available. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. The *filled bar* represents overall accuracy on all trials the monkeys chose to take. The *unfilled bar* represents overall accuracy on all trials the monkeys were forced to take. Error bars represent ±1 SEM.

Use of the *decline-test* response differed as a function of difficulty level ($F_{4,20}$ =6.59, p<.01). Monkeys declined more difficulty level 5 trials than difficulty level 1, but this effect was not statistically significant (T_5 = -2.42, p=.06).

Overall performance on chosen tests was higher than performance on forced tests, but this difference was not statistically significant (T_5 =-1.89, p=.12).

Monkeys did not transfer use of the *decline-test* response as well for the prospective memory task as for the concurrent memory task. This could be related to the novel placement of the choice phase, as monkeys had more practice making metacognitive judgments concurrently with the test during Experiments 1-4. Another possibility is that the visibility of the test at concurrent choices provided additional information, such as familiarity, that made metacognitive choices easier.

11. General discussion

We used a series of perceptual and memory tests to evaluate whether monkeys monitor private and public cognitive states to make metacognitive judgments. On the initial training task, eight of 12 monkeys learned to use the *decline-test* response adaptively, demonstrating a 30% difference between use of the *decline-test* response between difficulty level 1 and 5 trials. On subsequent generalization tests that used new perceptual discriminations, group performance was initially poor. Though some individuals did show robust transfer, the group on average used the *decline-test* response much more often than they had during the preceding discrimination. When use of the *decline-test* response is at ceiling, it is difficult to detect performance differences and this may partially account for poor transfer. Transfer of the *decline-test* response did improve with additional rounds of generalization testing. By the third generalization experience (Experiment 4), monkeys succeeded in the first transfer session, reflected in the significant difference in frequency of use of the *decline-test* response between the easiest and hardest trials. It is notable that this successful generalization followed experience using the *decline-test* response in multiple perceptual discrimination contexts. Monkeys subsequently generalized this behavior to a memory test, which, like the perceptual tasks, allowed them to make judgments concurrently, while the test was visible. This successful generalization of metacognitive responding to a small set matching-to-sample test is significant because, unlike the perceptual tests that preceded it, the display in the memory tests contains the same elements on every trial. Thus it is not possible for the specific stimuli present at test to occasion use of the *decline-test* response. This makes it more likely that the cue controlling use of the *decline-test* response is a private, internal representation. Monkeys were also moderately successful in making prospective metacognitive judgments. Together, these results provide preliminary evidence that a common private assessment of cognitive state, relevant to both memory and perceptual discrimination, cues use of the *decline-test* response.

The possibility that monkeys make metacognitive judgments on the basis of a general private cue raises questions about the nature of this private cue and the mechanism of generalization among tasks. Memory strength has been proposed as a basis for accurate metacognitive responding in memory tests (e.g., Hampton, 2001) and could be used in Experiments 5 and 6, but memory strength would not be relevant in Experiments 1-4, which used perceptual discriminations. Either different cues control the *decline-test* response in different tasks, or some more general “difficulty signal” controls behavior across different tasks. Thus, generalization may reflect a state akin to “uncertainty,” as put forth previously by Smith and colleagues (e.g., Smith et al., 2012).

Although a general cue like knowledge state would be available across all tasks, monkeys transferred poorly in Experiments 2 and 3. A possible explanation for this seeming inconsistency relies on the relation between public and private cues. It seems likely that public and private cues complement and inform one another. For example, metacognitive judgments could initially be based on a more overt public cue, such as the specific stimuli present at test. This public cue would likely be more salient than comparatively subtle assessments of private knowledge states. By this explanation, a private cue could initially be present, but too subtle to attend to in making metacognitive discriminations. With exposure to multiple different tests, the association between the *decline-test* response and specific stimuli would be degraded while the association with private cues would strengthen. After the *decline-test* response had been presented across several tasks, the memorization of multiple public cues would become an inefficient way to complete metacognitive judgments. Instead, a more abstract cue, such as a private judgment of "uncertainty" would become salient because it is common across tasks. This account predicts weak initial generalization and also allows for generalization from perceptual to memory tests.

An alternative possibility is that use of the *decline-test* response was controlled by a public cue that was available across tasks. We attempted to reduce the probability that public cues controlled metacognitive behavior by using very different displays in each perceptual generalization task. In Experiment 5 and Experiment 6, stimulus displays contained the same four images every trial, which further diminishes the probability that some aspect of test appearance could control behavior. Although delay length in Experiments 5 and 6 could provide a public cue that predicts test performance, this cue is not available in the prior tasks, and is thus unlikely to provide a basis for rapid generalization. Still, subjects likely displayed behaviors such as hesitation or vacillation while choosing among test items, which could provide salient public cues about the likelihood of success on a given trial. For this reason, the manipulation in Experiment 6, presenting the metacognitive choice prospectively, is particularly important for dissociating the use of public and private cues. Because subjects made metacognitive judgments prior to the test, behavioral responses to the appearance of the test could not guide use of the *decline-test* response.

On the first generalization session in Experiment 2, two monkeys chose the *decline-test* response for every trial on which it was available and some other monkeys chose the *decline-test* response with high frequency. This pattern of responding resulted in a higher overall proportion of trials declined in Experiment 2 compared with training performance in Experiment 1. Several possibilities could account for this result. First, it is possible that this result reflects a general state of confusion. In such a case, high use of the *decline-test* response could reflect a rule consistent with evaluation of a private state, wherein general uncertainty about the test elicits use of the *decline-test* response. Because metacognitive choice reflects a subjective state, it can be partially dissociated from objective difficulty. Forced test performance was initially titrated to a similar range of accuracy for each task, but it is possible that individual monkeys perceived the tasks to be of different subjective difficulty. More novel tasks, with which the monkeys have less practice, might be perceived as particularly difficult. This is consistent with some human models of metacognition, which posit that just as cognition is not entirely accurate, metacognition is also subject to errors (Nelson, 1996). We cannot evaluate subjective sense of difficulty. A final legitimate possibility is that the animals simply did not know how to make appropriate use of the response, and defaulted to a single response. However, no animals defaulted to always taking the test. It is possible that monkeys preferred a consistent but marginally lower rate of reinforcement. Thus, this pattern of performance neither provides strong evidence for or against control of the *decline-test* response by private cues. In any case, it appears that monkeys established a very low threshold for use of the *decline-test* response, as evidenced by the fact that use on the easiest trials never fell to 0% for any monkey on any transfer task.

We used this set of experiments to assess the extent to which public or private cues control rhesus monkeys' metacognitive choices. We tested a large number of subjects across five consecutive generalization tasks, which spanned multiple domains. Therefore use of the *decline-test* response could not be controlled by any obvious task-specific public cue. The results presented here provide provisional evidence that rhesus monkeys may be able to use a domain-general, private cue such as "uncertainty" to monitor that status of cognitive processes and knowledge states, as has been proposed by Smith and colleagues (e.g., Smith et al., 2012). Thus, these data lay promising ground work for future studies to explore the types of information that are available to monitoring, the conditions under which metacognition can take place, and the neurobiological foundations of this behavior.

12. References

Basile, B. M., & Hampton, R. R. (in press). Metacognition as discrimination: Commentary on Smith et al. (2013). . *Journal of Comparative Psychology*.

Basile, B. M., Hampton, R. R., Suomi, S. J., & Murray, E. A. (2009). An assessment of memory awareness in tufted capuchin monkeys (Cebus apella). *Animal Cognition, 12*(1), 169-180.

Brauer, J., Call, J., & Tomasello, M. (2004). Visual perspective taking in dogs (Canis familiaris) in the presence of barriers. [Article]. *Applied Animal Behaviour Science, 88*(3-4), 299-317. doi: 10.1016/j.applanim.2004.03.004

Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition, 4*, 207-220.

Foote, A. L., & Crystal, J. D. (2007). Metacognition in the rat. *Current Biology, 17*(6), 551-555.

Hampton, R. R. (2001). Rhesus monkeys know when they remember. [Article]. *Proceedings of the National Academy of Sciences of the United States of America, 98*(9), 5359-5362. doi: 10.1073/pnas.071600998

Hampton, R. R. (2009a). Focusing the uncertainty about nonhuman metacognition. . *Comparative Cognition & Behavior Reviews, 4*, 56-57.

Hampton, R. R. (2009b). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition & Behavior Reviews, 4*, 17.

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*, 239-254.

Inman, A., & Shettleworth, S. J. (1999). Detecting metamemory in nonverbal subjects: A test with pigeons. *Journal of Experimental Psychology-Animal Behavior Processes, 25*(3), 389-395.

Kornell, N. (2013). Where is the "meta" in animal metacognition? *Journal of Comparative Psychology*.

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

Metcalfe, J. (2008). Evolution of metacognition. *Handbook of metamemory and memory*, 29-46.

Nelson, T. O. (1996). Consciousness and metacognition. [Article; Proceedings Paper]. *American Psychologist, 51*(2), 102-116. doi: 10.1037//0003-066x.51.2.102

Shettleworth, S. J., & Sutton, J. E. (2003). Animal metacognition? It's all in the methods. *Behavioral and Brain Sciences, 26*(3), 353-+.

Smith, J. D. (2009). The study of animal metacognition. [Review]. *Trends in Cognitive Sciences, 13*(9), 389-396. doi: 10.1016/j.tics.2009.06.009

Smith, J. D., Couchman, J. J., & Beran, M. J. (2012). The highs and lows of theoretical interpretation in animal-metacognition research. [Review]. *Philosophical Transactions of the Royal Society B-Biological Sciences, 367*(1594), 1297-1309. doi: 10.1098/rstb.2011.0366

Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Rhesus monkeys (Macaca mulatta) adaptively monitor uncertainty while multi-tasking. [Article]. *Animal Cognition, 13*(1), 93-101. doi: 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 BOTTLE-NOSED-DOLPHIN (TURSIOPS-TRUNCATUS). [Article]. *Journal of Experimental Psychology-General*, *124*(4), 391-408. doi: 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*(1), 75-97.

Templer, V. L., & Hampton, R. R. (2012). Rhesus monkeys (Macaca mulatta) show robust evidence for memory awareness across multiple generalization tests. [Article]. *Animal Cognition, 15*(3), 409-419. doi: 10.1007/s10071-011-0468-4

Washburn, D. A., Gulledge, J. P., Beran, M. J., & Smith, J. D. (2010). With his memory magnetically erased, a monkey knows he is uncertain. [Article]. *Biology Letters, 6*(2), 160-162. doi: 10.1098/rsbl.2009.0737

Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys (Macaca mulatta) immediately generalize the uncertain response. [Article]. *Journal of Experimental Psychology-Animal Behavior Processes, 32*(2), 185-189. doi: 10.1037/0097-7403.32.2.185