ELSEVIER

An assessment of domain-general metacognitive responding in rhesus monkeys

Emily Kathryn Brown [a],[*], Victoria L. Templer [b], Robert R. Hampton [a]

[a] Department of Psychology and Yerkes National Primate Research Center, Emory University, 201 Downman Dr., Atlanta, GA 30322, USA
[b] Providence College, 1 Cunningham Sq., Providence, RI 02918, USA

ABSTRACT

Metacognition is the ability to monitor and control one's cognition. Monitoring may involve either public cues or introspection of private cognitive states. We tested rhesus monkeys (*Macaca mulatta*) in a series of generalization tests to determine which type of cues control metacognition. In Experiment 1, monkeys learned a perceptual discrimination in which a "*decline-test*" response allowed them to avoid tests and receive a guaranteed small reward. Monkeys declined more difficult than easy tests. In Experiments 2–4, we evaluated whether monkeys generalized this metacognitive responding to new perceptual tests. Monkeys showed a trend toward generalization in Experiments 2 & 3, and reliable generalization in Experiment 4. In Experiments 5 & 6, we presented the *decline-test* response in a delayed matching-to-sample task. Memory tests differed from perceptual tests in that the appearance of the test display could not control metacognitive responding. In Experiment 6, monkeys made prospective metamemory judgments before seeing the tests. Generalization across perceptual tests with different visual properties and mixed generalization from perceptual to memory tests provide provisional evidence that domain-general, private cues controlled metacognition in some monkeys. We observed individual differences in generalization, suggesting that monkeys differ in use of public and private metacognitive cues.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Imagine that a child asks if I can spell a word on their spelling list. Any well-informed observer could predict my performance about as accurately as I could, based on public cues, like my age, education, or experience. In contrast, if my friend asked me to spell a specific familiar but challenging word, like "bureaucracy," I might have to try to remember the spelling to gauge my ability to answer correctly. Although public cues, such as my latency to respond, could contribute to my judgement, the addition of the private result of my memory search would improve the accuracy of my judgment over that of an observer. Metacognition, or thinking about thinking, can be controlled by multiple cues, both public and private (Flavell, 1979; Hampton, 2009a; Basile and Hampton, 2014). Although both public and private cues can serve similar functions in improving the efficiency of cognition, the use of introspective private cues may be of particular theoretical interest to the extent that use of such cues implies explicit cognition, self-awareness, or consciousness. The extent to which nonhuman metacognition results from pub-

lic and private cues remains undetermined (e.g., Hampton, 2009b; Jozefowiez et al., 2009; Kornell, 2013; Le Pelley, 2012; Smith et al., 2014).

Many metacognition paradigms for nonhumans have used psychophysical discriminations as the primary task. The discriminanda differ along a continuum, for example, sparse vs. dense fields of dots, or high vs. low tones, with some difficult trials that fall close to the just noticeable difference. Rats, humans, a dolphin, birds, and monkeys have, at least under some circumstances, selectively avoided difficult psychophysical discriminations when given the option (Foote and Crystal, 2007; Smith et al., 1995, 1997; Nakamura et al., 2011). Other paradigms assess the ability of subjects to make metacognitive judgments about memory or learning, by testing whether subjects decline memory tests when they have forgotten or when they have not mastered a task (Fujita, 2009; Hampton, 2001; Kornell et al., 2007; Morgan et al., 2014; Suda-King, 2008; Suda-King et al., 2013; Templer and Hampton, 2012; Washburn et al., 2010). A major aim of metacognition research in nonhumans has been to distinguish performance controlled by private cues from performance controlled by public cues (Hampton, 2009b; Roberts et al., 2012; Smith, 2009; Smith et al., 2012). Because introspective private cues cannot be directly manipulated and observed by experimenters, the use of these cues can only be inferred by

* Corresponding author at: PAIS Building, 36 Eagle Row, Atlanta, GA 30322, USA.
E-mail address: emily.brown@emory.edu (E.K. Brown).

exclusion of public cues. If there appears to be no public cue that can account for metacognitive performance, private cues may be involved (Shettleworth and Sutton, 2003).

Generalization tests evaluate the extent to which metacognition depends on domain-general cues, by determining whether metacognitive responses acquired in one context generalize to new conditions (Basile et al., 2015; Hampton, 2001; Malassis et al., 2015; Smith et al., 2010; Templer and Hampton, 2012; Washburn et al., 2006). When an appropriate novel primary task is introduced, many cues that could have guided apparently metacognitive responding on the previous task are eliminated, including those based on specific public properties of the stimulus display at test. If metacognitive judgments are controlled by domain-general cues such as confidence or uncertainty, metacognitive performance should immediately transfer to any new primary cognitive test that elicits similar private cognitive states. In contrast, if metacognitive judgments are controlled by public cues that are specific to particular tasks, generalization among tasks will not occur. When the public cues that previously controlled metacognitive responding become unavailable, subjects would lose the basis for metacognitive judgments. Generalization of metacognitive performance across a variety of primary cognitive tests would therefore provide evidence for control by domain-general cues. However, we emphasize that while private cues such as states of uncertainty or confidence would likely be domain general, the observation of generalization among primary cognitive tasks does not necessitate control of metacognitive responding by private cues. Some public cues could also allow for substantial generalization. For example, subjects might attend to their own hesitation or vacillation, avoiding tests if response latencies are long (e.g., Hampton, 2009b).

Memory tests may provide a more stringent assessment of whether private cues control metacognition than do perceptual tests. Unlike psychophysical tasks, in memory tests trial difficulty is not determined by the appearance of the test display alone. Performance on memory tests depends on the persistence of an internal representation of the to-be-remembered sample over a delay period (Metcalfe, 2008). Given that subjects are sometimes correct and sometimes wrong in memory tests, the quality of the internal representation of the sample clearly varies among trials. Primates have often succeeded at making metacognitive judgments about memory performance (Basile et al., 2015; Hampton, 2001; Kornell et al., 2007; Templer and Hampton, 2012; Washburn et al., 2010), whereas other nonhumans have shown more mixed results (Adams and Santi, 2011; Brauer et al., 2004; Goto and Watanabe, 2012; Inman and Shettleworth, 1999; Iwasaki et al., 2013; Roberts et al., 2009). Even when memory tests are used, a subject's own publicly observable behavior such as vacillation or hesitation could control metacognitive judgments under some conditions (Hampton, 2009a,b).

Memory tests provide unique opportunities to assess control of metacognition by private cues because the metacognitive choice can be provided before the test. Performance in memory tests depends on a representation of the sample over a delay period, and we can expect the quality of that representation to vary across trials and over delays. With psychophysical tasks, subjects cannot possibly judge the difficulty of a trial prior to presentation of the test array. In contrast, there is at least the potential for subjects to monitor the quality of a memory trace during the delay interval of a memory test, a time during which the test display is not visible and vacillation and hesitation cannot occur. Public cues such as vacillation or hesitation can therefore be eliminated by requiring subjects to make prospective metacognitive judgments, before test stimuli are seen (Hampton, 2001).

It is likely that a variety of cues can control metacognitive responding in both humans and nonhumans and that subjects use different cues depending on what is salient and reliable in a given situation. But it is also possible that some domain-general state like "uncertainty" or "confidence" that could be modulated in a wide variety of tasks is the critical cue for a wide range of metacognitive judgments (Smith et al., 2012, 2014). Weak memories, difficult discriminations, and incomplete learning could all lead to changes in a domain-general state of "confidence." One challenge to determining which cues control metacognitive performance is that different paradigms have only rarely been directly compared with generalization tests (Washburn et al., 2006; Kornell et al., 2007). Thus, even if private cues appear to control metacognitive responding in a single prospective metamemory judgment, it is not clear that the same private cue would be responsible for metacogntive performance in a psychophysical task. For example, it has been proposed that "memory strength" might control metacognitive responding in prospective metamemory judgments (Basile et al., 2015; Hampton, 2001; Templer and Hampton, 2012), but memory strength would be irrelevant in psychophysical tasks.

We evaluated the bases of metacognitive responding in monkeys by implementing a variety of transfer tasks that differed in the availability of specific public and private cues for metacognitive judgments. We assessed immediate generalization of the metacognitive response in each new task. 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 control of the *decline-test* response generalized in the absence of stimulus-specific cues from initial training. With Experiment 5, we evaluated whether the same cues control metacognitive responding in psychophysical and memory tests, by transferring monkeys to a delayed matching-to-sample task. In Experiment 6, we evaluated whether the cue controlling metacognition was private with a final transfer to prospective metamemory judgments.

If metacognitive responding is under the control of private cues, such as "confidence," monkeys should immediately transfer use of the *decline-test* response across the perceptual tests and the concurrent and prospective metamemory judgments. If monkeys' judgments are instead controlled by public cues, such as their own behavior or some aspect of the stimulus display at test, they should fail to generalize use of the *decline-test* response during the series of transfer tests.

2. Materials and methods

2.1. 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, in which monkeys learned to use a *decline-test* option in the context of a delayed match-to-location paradigm using an apparatus like a WGTA (Templer and Hampton, 2012).

2.2. Apparatus

We tested monkeys in their home cages, using portable touchscreen 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 mg banana-flavored pellets (Bio-Serv, Frenchtown, NJ) or 97 mg fruity-flavored nutritionally complete primate pellets (Purina TestDiet, Richmond, IN).

Most monkeys received banana-flavored pellets. Monkeys that worked especially slowly were given fruity-flavored pellets to enhance their motivation. We presented stimuli and collected responses using programs written in Presentation (Neurobehavioral Systems, Albany, CA).

2.3. Procedure

2.3.1. Monkey housing and testing conditions

During testing, pair-housed 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. Concurrent experience included tests of memory for order, and was not expected to affect performance in the present tasks. 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.

2.3.2. 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 with the touchscreen as responses. 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 included as the hardest trials. On such unsolvable trials, one location was still assigned as the "target", and selection of this location resulted in reward as on correct trials; this target location was pseudo-randomly assigned to ensure that each of the locations was used equally often. Responses on unsolvable trials were rewarded at the chance rate. Therefore, difficulty level 5 trials served to anchor performance by guaranteeing that the most challenging trials be sufficiently difficult to elicit use of the *decline-test* response.

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. Initially, all timeouts following incorrect responses were 500-milliseconds, but during training, the timeout was titrated for each monkey to increase the salience of trial difficulty. 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-milliseconds to 240-s, according to individual learning and motivation to use the *decline-test* response.

A 2-s 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.
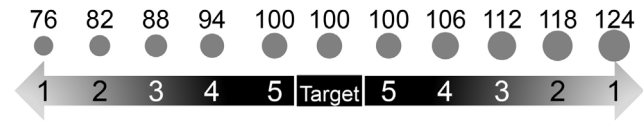

Fig. 1. Stimuli used in the final phase of the size discrimination (top). Labels on the circles indicate actual diameters, in pixels, used in the experiment, but were not shown to the monkeys. At each level of difficulty (bottom), there were two absolute distracter sizes. The easiest distracters (i.e., 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.

3. Experiment 1- size discrimination

3.1. Training on size discrimination

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 (Fig. 1).

We titrated the difficulty of the task by changing the dimensions, and thus the discriminability of the test stimuli. Training consisted of 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.

3.2. Training on the decline-test response

Trials proceeded as described above, except for the addition of a metacognitive choice phase that allowed monkeys to take the test for a large reward if correct, or avoid the test for a small, guaranteed reward. In the metacognitive choice phase, two additional black and white clipart choice stimuli could be displayed concurrently with test stimuli. The *accept-test* stimulus, a check-marked square, was vertically centered on the right side of the screen. Touches to the *accept-test* stimulus extinguished metacognitive choice stimuli and made the choice stimuli responsive to touch. 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, was vertically centered on the left side of the screen. Selection of the *decline-test* stimulus resulted in the immediate presentation of a red bar at the top center of the screen. Touches to this *guaranteed small reward* stimulus resulted in a distinctive auditory signal and one food pellet.

On 2/3 of trials, subjects were presented with both metacognitive choice stimuli (Fig. 2, left side). On the other 1/3 of trials, only the *accept-test* stimulus was presented, forcing subjects to take the test (Fig. 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, such that each difficulty level was represented twice in every ten trials and each target location was correct twice every eight trials.

3.3. Data analysis

All proportions were arcsine transformed before statistical analysis to better approximate the normality assumption

Table 1
Distracter size, diameters in pixels.

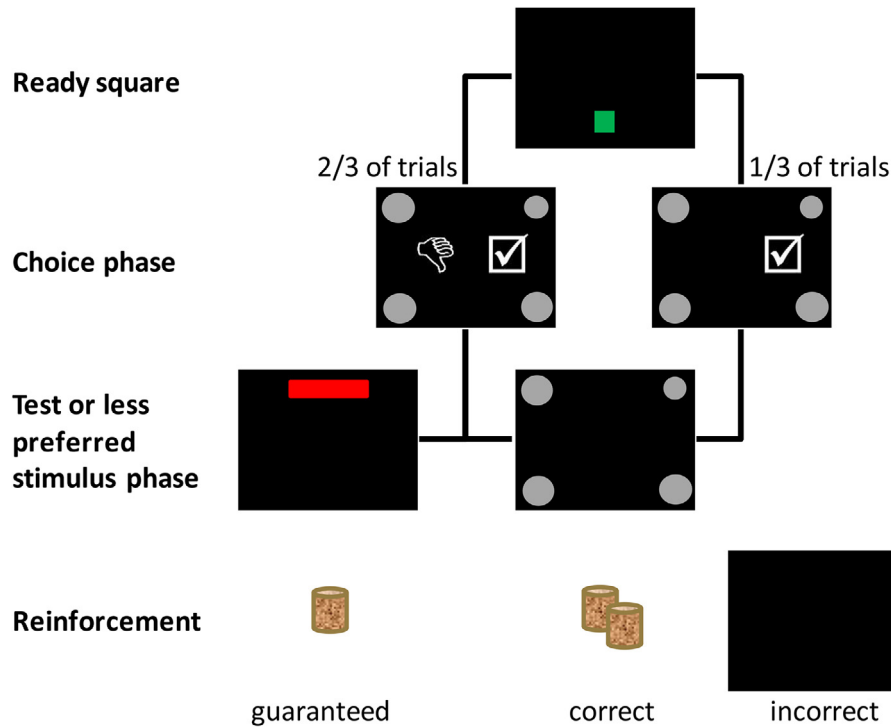|	easy				hard	target	hard				easy
Phase 1	50	54	58	62	66	100	134	138	142	146	150
Phase 2	60	68	76	84	92	100	108	116	124	132	140
Phase 3	76	82	88	94	100	100	100	106	112	118	124

Fig. 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. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

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 accuracy and proportion of trials on which monkeys used the *decline-test* response as a function 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. We also assessed differences in accuracy between forced and chosen tests, pooled across all levels of difficulty, using paired *t*-tests. Because all paired *t*-tests were planned *a priori*, we did not apply corrections to control for family-wise error. In Experiment 1, we conducted these statistical analyses on the final criterion session, to ensure that we had successfully trained initial use of the *decline-test* response. Thereafter, all analyses were conducted on the first session of a new generalization task, before monkeys had completed any extended training necessary to move on to a subsequent task.

3.4. 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 met the *decline-test* criterion of 30% difference in use of the *decline-test*

response on easiest and hardest trials. Because subsequent tasks evaluated transfer of criterion performance acquired in Experiment 1, these four were not included in this analysis or the remainder of this study.

Accuracy on forced test trials differed as a function of the similarity between the size of the distracters and the target (Fig. 3; $F(4, 28) = 61.94$, $P < 0.001$). Monkeys were significantly more accurate on difficulty level 1 trials than on difficulty level 5 trials ($t(7) = 13.41$, $P < 0.001$).

For the 8 monkeys that reached criterion for use of the *decline-test* response, use of the *decline-test* response differed as a function of task difficulty ($F(1.415, 9.90) = 17.80$, $P = 0.001$). Monkeys declined significantly more trials from difficulty level 5 than level 1 ($t(7) = -6.25$, $P < 0.001$).

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. Thus, monkeys could increase the proportion of trials resulting in reinforcement by declining mostly trials from higher difficulty-level discriminations, without improving accuracy on chosen over forced discriminations of the same difficulty.
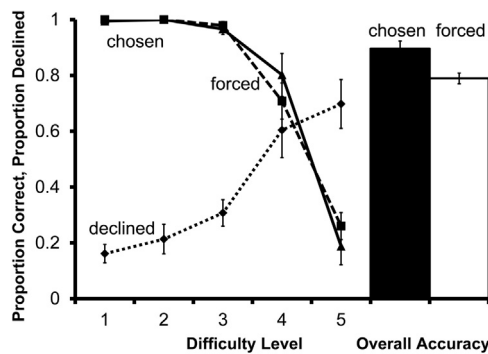
Fig. 3. Performance of 8 monkeys on the final criterion session of the size discrimination metacognition training task in Experiment 1. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which 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.

Overall accuracy on chosen tests was significantly higher than on forced tests ($t(7) = 3.23$, $P = 0.014$); thus monkeys experienced more reinforced trials by using the decline-test response. However, forced and chosen accuracy functions did not differ from one another ($F(1,7) = 0.08$, $P = 0.790$), indicating that monkeys discriminated among the different difficulty levels, but either did not perceive differences in difficulty between specific trials at a given difficulty level, or no such differences existed. If monkeys were able to detect differences in difficulty within a given difficulty level, we might expect that they would selectively avoid those trials, resulting in higher chosen than forced accuracy at a given level. For this task, experimenter-defined difficulty levels correspond to absolute differences in perceptibility, and all trials within a difficulty level contain identical elements. Within a difficulty level, there is little to cue which trials monkeys will get right and which trials they will get wrong. Cues associated with success or failure should be more salient between trials of different difficulty levels.

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, some specific aspect of the display could have controlled metacognitive responding. For example, the distance between stimuli and the edge of the screen, the overall luminance of the display, or some other feature could have cued the probability of reinforcement. To address the concern that visual cues present in the test could control metacognitive responding, we conducted a transfer test with substantially different stimulus appearance in Experiment 2.

4. Experiment 2- brightness discrimination

4.1. Rationale

Monkeys received extensive training with the *decline-test* response in the context of the size discrimination used in Experiment 1. It is possible that use of the *decline-test* response was controlled by learned associations between specific screen displays and probability of reinforcement. Generalization tests provide a means to assess which cues control metacognitive judgments because changing the primary task eliminates some public cues specific to the original task. If domain-general cues, such as private states of uncertainty, controlled use of the *decline-test* response in Experiment 1, monkeys should immediately transfer to a novel brightness discrimination in Experiment 2. This is because the new task should elicit similar private, cognitive states. In contrast, if

Fig. 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.

public cues specific to this particular test controlled the pattern of performance in Experiment 1, the monkeys should fail to rapidly generalize the metacognitive response.

4.2. Subjects

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

4.3. 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 (Fig. 4; Table 2).

4.4. 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 brightness discrimination had sufficient variation in difficulty to elicit both *decline-test* and *accept-test* responses. Note that this pre-transfer review does not involve any exposure to the *decline-test* response in the context of the new brightness discrimination.

4.5. Transfer of the decline-test response

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

4.6. Data analysis

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

4.7. Results and discussion

Accuracy on forced test trials differed as a function of the similarity between the brightness of the distracters and the target (Fig. 5; $F(4, 28) = 52.01$, $P < 0.001$); monkeys were significantly more accurate on the difficulty level 1 trials than on difficulty level 5 trials ($t(7) = 12.58$, $P < 0.001$).

In the first transfer session, use of the *decline-test* response showed a trend toward a relation with task difficulty, but the effect

Table 2
Distracter brightness in RGB values.

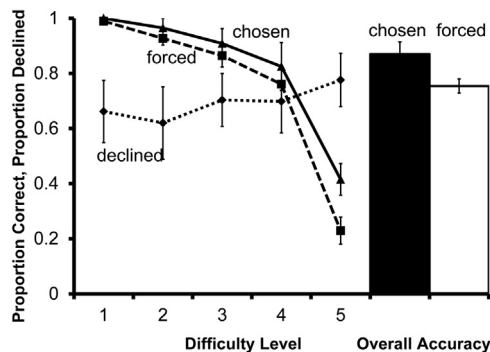|	easy				hard	target	hard				easy
Phase 1	64	74	84	94	104	128	152	162	172	182	192
Phase 2	64	74	84	94	128	128	128	162	172	182	192

Fig. 5. Performance of 8 monkeys on the brightness discrimination transfer test in Experiment 2. *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.

was not statistically significant (F(1.65,11.53) = 3.26, P = 0.082). 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 = 0.129). Thus, as a group, monkeys failed to generalize use of the metacognitive response to a new perceptual discrimination, suggesting that metacognitive responding in Experiment 1 was controlled by cues specific to that task. However, overall use of the *decline-test* response was high in Experiment 2 compared to Experiment 1 (69.17% (SEM = 0.106) of trials on which the response was available in Experiment 2, compared to 39.69% (SEM = 0.055) 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 = 0.021). 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 for the six monkeys for whom data were available (t(5) = 2.13, P = 0.086). This comparison does not include the two monkeys that declined all trials for which they had the option, because accuracy on chosen trials could not be determined.

As a group, the monkeys did not 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 several individual monkeys that did not use the decline-test response on 100% of opportunities appear to have generalized. Excessive use of the decline-test response may have been caused by the novelty of its appearance in a less familiar task. We reasoned that the *decline-test* response could be controlled by multiple cues, some specific to a given task, and some more general across tasks. Exposure to the *decline-test* response in multiple contexts may degrade the association between the *decline-test* response and specific test stimuli and strengthen the association with private cues that are similar across tests. Previous research in the domain of concept-learning has indicated that a greater number of training exemplars facili-

Fig. 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° longer or shorter than the target.

tates rule-learning (e.g., Wright and Katz, 2007). To provide further practice with the *decline-test* response in multiple contexts, we extended training and provided another novel generalization task, described in the next experiment.

5. Experiment 3- arc length discrimination

5.1. Rationale

The generalization test in Experiment 2 yielded ambiguous results. Some monkeys showed evidence of generalization; however, the effect of difficulty level on use of *decline-test* response was non-significant. 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 difficult to interpret. Given these results, and the possibility that additional generalization opportunities may decouple specific aspects of the test from the metacognitive response, we provided another transfer task, an arc length discrimination, in Experiment 3.

5.2. Subjects

Seven of the eight monkeys from Experiment 2 participated in Experiment 3. One monkey was released from our animal care and use protocols because he consistently finished experiments much more slowly than other monkeys. He had 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 the remainder of this study.

5.3. 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.

5.4. 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 (Fig. 6; Table 3A). 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

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

A

	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

B

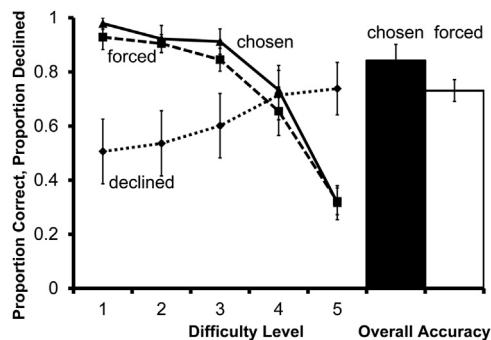|	easy				hard	target	hard				easy
Phase 1	80	75	70	65	60	45	30	25	20	15	10
Phase 2	80	75	70	65	45	45	45	25	20	15	10

Fig. 7. Performance of 7 monkeys on the arc length discrimination transfer test in Experiment 3. *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.

point, he was given distracters more discriminable from the target, to make the task easier (Table 3B).

5.5. 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 two 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.

5.6. Transfer of the decline-test response

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

5.7. Results and discussion

Accuracy on forced test trials differed as a function of the similarity between the arc length of the distracters and the target (Fig. 7; $F(4, 24) = 29.47$, $P < 0.001$); accuracy was higher on the easiest trials than unsolvable ones ($t(6) = 9.16$, $P < 0.001$).

Use of the *decline-test* response showed a trend toward a relation with task difficulty, but the effect was not statistically significant ($F(1.23, 7.40) = 4.59$, $P = 0.062$); however, the difference between difficulty level 5 vs. difficulty level 1 trials was significant ($t(6) = -2.50$, $P = 0.046$).

Average performance on chosen tests was higher than on forced, but this difference was not significant ($t(5) = 1.40$, $P = 0.221$). This comparison does not include the one monkey that declined all trials for which he had the option, because accuracy on chosen trials could not be determined. The monkey that declined all trials for which he had the option was one of the monkeys that did so in the previous experiment.

6. Experiment 4- rotation discrimination

6.1. Rationale

Experiment 3 provided stronger evidence of transfer than Experiment 2. 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. Although monkeys showed improved evidence of transfer, they still did not show the robust transfer that would be associated with strong control of metacognitive responding by a cue that is present across tasks. Given that the monkeys appeared to be increasing proficiency with the *decline-test* response, we presented monkeys yet another perceptual task: a rotation discrimination transfer test. This task could also provide new training exemplars that encourage attention to the commonalities shared across perceptual tasks.

6.2. Subjects

The seven monkeys from Experiment 3 participated in Experiment 4.

6.3. 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.

6.4. 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

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

Fig. 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° from the target, either clockwise or counter-clockwise.

the center point of the outer circle, but were identical along all other dimensions. The target stimulus was held at a constant rotation (Fig. 8; Table 4). 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 in Fig. 8.

6.5. 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. Because monkeys had extensive prior experience making metacognitive judgments in the context of known discriminations, we determined that it was not necessary to use as many review sessions as in previous experiments. 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 *decline-test* and *accept-test* responses.

6.6. 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.

6.7. Results and discussion

Accuracy on forced test trials differed as a function of the similarity between the degrees rotation of the distracters and the target (Fig. 9; $F(4, 24) = 23.99$, $P < 0.001$); accuracy was higher on difficulty level 1 than level 5 ($t(6) = 7.50$, $P < 0.001$).

Use of the *decline-test* response differed as a function of difficulty level ($F(1.432, 8.592) = 5.93$, $P = 0.031$), and monkeys declined significantly more difficulty level 5 trials than difficulty level 1 ($t(6) = −2.81$, $P = 0.031$). This was the first task on which mon-
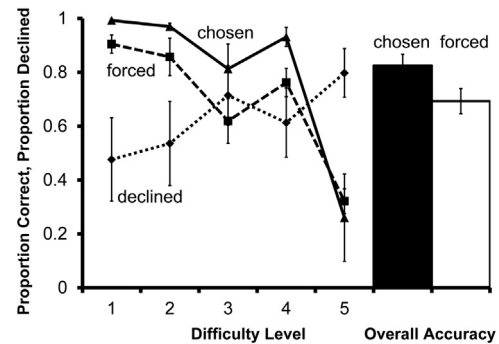

Fig. 9. Performance of 7 monkeys on the rotation discrimination transfer test in Experiment 4. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. 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.

keys showed generalization of the *decline-test* response as a group, indicated by the relation between use of the response and trial difficulty.

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 ($t(5) = 3.07$, $P = 0.028$), indicating that differential decline of difficult trials improved overall performance. This comparison does not include the one monkey that declined all trials for which he had the option, because accuracy on chosen trials could not be determined. The monkey that declined all trials for which he had the option had not done so in the previous experiments.

Monkeys transferred adaptive use of the *decline-test* response for this final perceptual task. We identify several possible reasons monkeys improved across subsequent generalization tasks. 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. Another possibility is that application of the *decline-test* response to multiple perceptual domains degraded task-specific associations and increased control by a general cue shared across tasks. Use of the *decline-test* response on this final perceptual task generalized immediately, suggesting that monkeys used a cue that was available across perceptual domains. It is possible that monkeys were able to generalize use of the *decline-test* response to new perceptual tasks because metacognitive responding was controlled by a domain-general private cue, such as "uncertainty." However, in each perceptual task, the appearance of the display likely provided a highly salient public cue about probability of success and reinforcement on the test, perhaps in the form of self-generated behavior. On difficult perceptual tests, for which stimuli were highly similar or even identical to one another, the objective perceptual similarity of the stimuli could prevent the monkey from immediately arriving at a solution. The absence of a decisive response, a public cue, could then control use of the *decline-test* response. Memory tests provide an opportunity to further reduce the possibility of control of the metacognitive response by display-specific public cues because memory tests can be conducted using identical test displays on every trial. If found, metacognitive responding in such tests is more likely to result from

private metacognitive cues, possibly an internal representation of the stimulus seen at study or degree of "uncertainty." To further evaluate the type of cues monkeys use to make the *decline-test* response the next transfer tests were to memory tasks.

7. Experiments 5 and 6- transfer to memory tests

7.1. General methods for memory tasks

We trained monkeys on a delayed matching-to-sample (DMTS) memory task to assess generalization of the *decline-test* response to a non-perceptual cognitive domain. A small set of 4 clipart images was used across all sessions, such that every image was seen at test on every trial. All responses required two touches (FR2) to prevent recording undirected contacts with the touchscreen as responses. To start a trial, monkeys touched a green ready square at the bottom center of the screen. A sample image then appeared in the center of the screen. Touches to this image resulted in a blank screen for a delay. Choice of the sample image seen immediately 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-s 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 (Fig. 10).

8. Experiment 5- concurrent choice metamemory

8.1. Rationale

The fact that monkeys generalized immediately to the final perceptual discrimination in Experiment 4 suggests that by this point in training, domain-general cues had gained control of the metacognitive response. But because the task difficulty in Experiments 1–4 relies on the appearance specific to the stimuli at test, it is still possible that some public feature of the test displays, shared at least by the tests used in Experiments 3 and 4, controlled metacognitive responding. In Experiment 5 we used a memory task, with the same four images shown on every test. Because all tests used the same four images, the appearance of the test display cannot provide any useful cues about task difficulty. If use of the metacognitive response in Experiment 4 was controlled by a domain-general cue, we expect monkeys to transfer use of the *decline-test* response to the memory domain. Generalization would not require control of the metacognitive response by a private cue. Public self-generated cues, such as vacillation or hesitation, could still control metacognitive responding in a memory test, but other public cues present across the perceptual discriminations are unlikely to be present in this novel mnemonic test.

8.2. Subjects

The seven monkeys from Experiment 4 participated in Experiment 5.

8.3. 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.

8.4. Training for memory task

Prior to Experiment 5, monkeys completed 20 sessions of DMTS with retention intervals of 6, 12, 24, and 48-s delays intermixed. After these training sessions, a very short delay length was added, so that delays lasted 0.2, 6, 12, 24, or 48-s. Difficulty level for this task was based on delay length, such that difficulty level 1 trials included a 0.2-s delay, difficulty level 2 trials included a 6-s delay, etc.

8.5. 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 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, the ITI was increased by 5-s to decrease interference. Performance was re-evaluated every 4 sessions, at which time ITI was increased by 5-s intervals or monkeys were returned to pre-transfer review.

8.6. 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 a changing FR to obtain his small guaranteed reward, as described in Experiment 4. 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. As in prior experiments, in Experiment 5, the choice stimuli were presented concurrently, at the same time as test stimuli.

8.7. Results and discussion

Accuracy on forced test trials differed as a function of the delay length (Fig. 11; $F(4, 24) = 3.12$, $P = 0.034$); accuracy was higher on difficulty level 1 than level 5 ($t(6) = 2.45$, $P = 0.050$).

Use of the *decline-test* response differed as a function of task difficulty ($F(4, 24) = 23.94$, $P < 0.001$), and monkeys declined significantly more difficulty level 5 trials than difficulty level 1 ($t(6) = -7.77$, $P < 0.001$).

Monkeys were numerically more accurate on forced tests than on chosen tests on difficulty level 5 trials. Because monkeys were declining the vast majority of difficulty level 5 trials, there are likely too few chosen trials represented for such a forced-chosen accuracy difference to be meaningful. Overall performance on
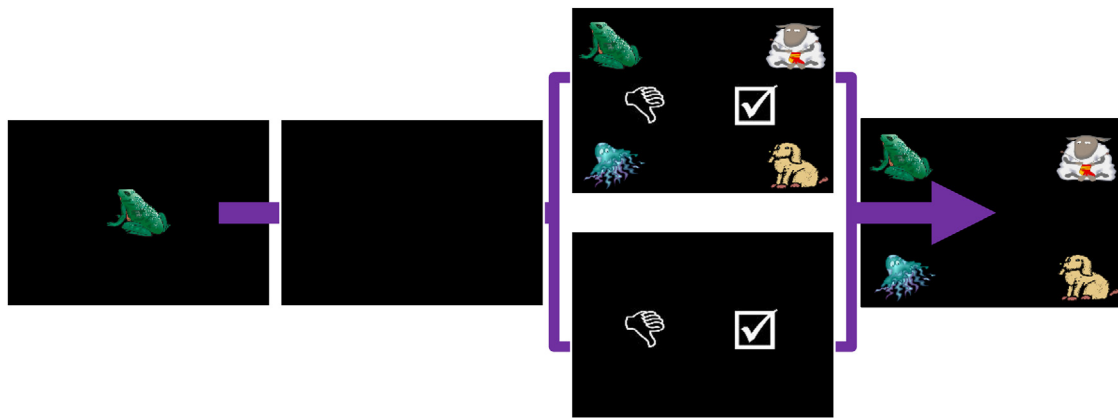
Fig. 10. Steps to complete a trial of the memory task with metacognitive choice stimuli. Monkeys touched the green ready square to initiate trials (not shown). A sample clipart image then appeared on screen. On 2/3 of trials, the *decline-test* and *accept-test* stimuli appeared after a delay. The choice stimuli could appear with the test of memory in *concurrent* metamemory judgments (Experiment 5, above) or before the test in the case of *prospective* metamemory judgments (Experiment 6, below). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
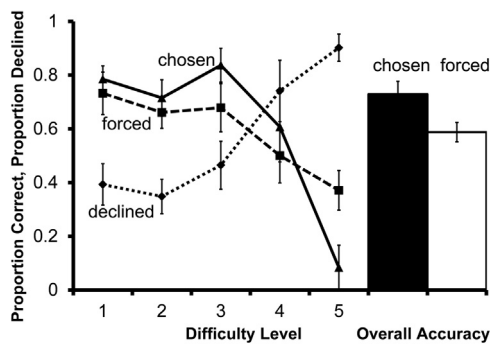

Fig. 11. Performance of 7 monkeys on the concurrent metamemory transfer test in Experiment 5. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. 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.

chosen tests trended toward being higher than performance on forced tests, although the difference was not statistically significant ($t(6) = -2.29$, $P = 0.062$).

Monkeys transferred adaptive use of the *decline-test* response from familiar perceptual tests to the novel memory test. This immediate generalization indicates that the use of the *decline-test* response was controlled by a cue shared across both perceptual discriminations and memory tests. Transfer could indicate control of metacognitive responding by a private cue, or control by a self-generated public cue shared across the perceptual discriminations and memory test. Monkeys are most likely to self-generate public cues in response to the appearance of the test, so in Experiment 6, we decoupled the metacognitive decision from the appearance of the test. By presenting the prospective metamemory judgments before the test stimuli appeared, we tested whether metacognitive responding is controlled by a self-generated public cue such as hesitation, or by a private cue shared across tests, such as a state of "uncertainty." Generalization to prospective metamemory judgments would be most likely if private cues contribute to metacognitive responding.

9. Experiment 6- prospective metamemory

9.1. Rationale

Prospective metamemory judgments allow presentation of the metacognitive choice before the test can elicit public responses such as hesitation or vacillation. Thus prospective metamemory judgments allow us to discriminate between control of the metacognitive response by self-generated public cues and by private cues.

9.2. Subjects

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

9.3. 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.

9.4. Pre-transfer review

Prior to the transfer task, monkeys were required to complete at least 5 sessions of prospective metamemory judgments 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.

9.5. Transfer of the decline-test response

In Experiment 6, the choice stimuli were presented prospectively before test stimuli.
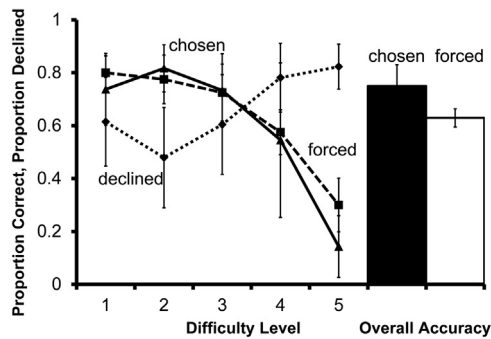
Fig. 12. Performance of 6 monkeys on the prospective metamemory transfer test in Experiment 6. *Solid* and *dashed lines* indicate accuracy on chosen and forced tests according to difficulty level. The *dotted line* indicates proportion of choice trials for which the *decline-test* response was used. 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.

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.

9.6. Results and discussion

Accuracy on forced test trials differed as a function of the delay length (Fig. 12; $F(4,20) = 6.70$, $P = 0.001$); accuracy was higher on difficulty level 1 than level 5 ($t(5) = 4.37$, $P = 0.007$).

Use of the *decline-test* response differed as a function of difficulty level ($F(4, 20) = 6.59$, $P = 0.001$). Monkeys trended toward declining more difficulty level 5 trials than difficulty level 1, but this effect was not statistically significant ($t(5) = -2.42$, $P = 0.060$).

Monkeys declined numerically more difficulty level 1 than difficulty level 2 trials. Higher use of the *decline-test* response on difficulty level 1 trials was matched by a numerical dip in proportion correct on chosen difficulty level 1 trials compared with difficulty level 2 trials. Although we might expect that monkeys would decline the fewest difficulty level 1 trials, congruent choice accuracy and *decline-test* use may indicate that monkeys are attending to subjective trial difficulty.

Overall performance on chosen tests was higher than performance on forced tests, but this difference was not statistically significant ($t(5) = -1.89$, $P = 0.117$).

Monkeys used the *decline-test* response more on more difficult trials, indicating generalization of metacognitive responding to novel prospective metamemory judgments. However, they did not avoid the most difficult trials significantly more than the least difficult trials, as they did in Experiment 5, nor did their use of the *decline-test* response significantly improve chosen over forced accuracy. Generalization of metacognitive responding was weak in this last transfer test, but the general pattern of *decline-test* use across difficulty levels suggests that metacognitive behavior may be controlled, at least in part, by a domain-general private cue.

10. General discussion

Generalization of metacognitive performance between perceptual and memory tasks suggests that metacognitive responding was controlled, in part, by a domain-general, private cue, at least after extensive training on a series of diverse perceptual and mnemonic tasks as was done here. We used a relatively large number of primary cognitive tests, both perceptual and mnemonic, and a relatively large number of subjects (n = 12 initially) to evaluate the extent to which monkeys make adaptive metacognitive responses. Perhaps because our subject pool was so large—one of the largest reported for a study of this kind – we found considerable

inter-individual variation in acquisition of metacognitive responding and in generalization between tasks. Whereas some monkeys generalized metacognitive performance after training in a single perceptual task, others required training in additional tasks, or transferred inconsistently on a task-by-task basis. In Experiment 5, monkeys transferred adaptive metacognitive responding from the perceptual to the memory domain, suggesting that monkeys relied on a domain-general cue. In Experiment 6, monkeys made prospective metamemory judgments. Because these judgements are made prior to the appearance of the test, they cannot be controlled by public cues present at test. Together, transfer in Experiment 5 and partial transfer in Experiment 6 suggest that metacognitive responding was controlled, at least in part, by a domain-general private cue, such as "uncertainty."

Because each generalization test involved substantial change in the specific stimuli present at test, the specific test stimuli could not occasion use of the *decline-test* response across tasks. In the perceptual tests, the appearance of the display varied substantially across tasks. Because target stimuli were not visually similar, the visual display is unlikely to have controlled generalization across perceptual domains. The generalization from perceptual metacognition to metamemory in Experiment 5 provides strong evidence against specific test stimuli controlling the decline test response. In memory tests, the stimulus displays contained the same four images every trial. Because the same four images were present on each trial, but which image was correct varied from trial to trial, specifics of the test display could not elicit adaptive use of the *decline-test* response. Perceptual tests and memory tests were procedurally and visually very distinctive, and the monkeys generalized between them here, as has been demonstrated in prior studies (Kornell et al., 2007; Washburn et al., 2006). However, in the case of the tasks used in prior studies, self-generated hesitation or vacillation in response to the test could have provided a salient cue to control generalized metacognitive responding. Prospective metamemory judgments, as we used here, attenuate the likelihood that public cues alone could control metacognitive responding. Because subjects made metacognitive judgments prior to the test, self-generated behavioral cues elicited at test, such as hesitation or vacillation, could not control use of the *decline-test* response. These results provide preliminary evidence that a common private assessment of cognitive state, relevant to both memory and perceptual discrimination, can control use of the *decline-test* response in monkeys.

10.1. Cues controlling the use of the decline test response

Because monkeys generalized use of the decline test response from perceptual tests to memory tests, the cue controlling use of this response appears to be a cognitive state elicited by both tasks. The strength of a memory for the sample seen at study has been proposed as a private cue that could be the basis for accurate metacognitive responding in memory tests (e.g., Hampton, 2001, 2009b). Such a cue could account for performance in Experiments 5 and 6, where memory was the relevant cognitive domain. However, memory strength would not be relevant to metacognitive responding to perceptual discrimination tests in Experiments 1–4, and therefore memory strength would not provide a basis for generalization between perceptual and memory tests. Either a general "difficulty signal" controlled behavior across tasks, or different cues controlled the *decline-test* response in our different tasks. We note that the cues controlling metacognitive responding in this study could be different than those in other studies. It is possible that monkeys trained exclusively in metamemory tasks do indeed attend to memory strength (Basile et al., 2015; Hampton, 2001; Templer and Hampton, 2012), but when trained in multiple tasks, as we did here, cues general to both perceptual and memory tasks gain control of the metacognitive response. Metacognitive transfer

might be considerably easier within a single cognitive domain, such as variants of memory tests (e.g., Basile et al., 2015), than between domains, such as the transfers between perceptual and memory tests done here.

10.2. Uncertainty

The generalization of metacognitive responding described here is consistent with "uncertainty" being the private cue controlling the metacognitive response, as put forth previously by Smith and colleagues (e.g., Smith et al., 2012). Nonhuman's subjective internal states are challenging to operationalize, so it is helpful to consider what such an uncertainty signal would entail. For example, uncertainty might be related to processing fluency.

Task fluency, the ease of processing the primary cognitive task, could control metacognitive responding (Kornell, 2013). High fluency would correspond to low uncertainty. Monkeys could attend to different types of task fluency across the two primary tasks: fluency in response to the test in Experiments 1–5 and ease of retrieval in Experiments 5–6. Alternatively, monkeys may rely on some sort of domain-general fluency available across tasks, such as the ease with which the target item is retained or brought into working memory. Kornell (2013) has proposed tests of a task fluency hypothesis, in which ease of processing could be manipulated by altering contrast in a perceptual task, or familiarity in a memory task, perhaps inducing metacognitive bias or errors independent of primary task accuracy. The individual differences in metacognitive responding that we observed in these studies may indicate individual differences in processing fluency, or perceived task difficulty.

10.3. Metacognitive responding may be controlled by both subjective and objective difficulty

Whereas we might expect that metacognitive responding controlled solely by a public cue would map rigidly onto objective task difficulty, some individuals immediately generalized to novel tasks, whereas others did not. This finding could be consistent with a subjective internal cue, such as processing fluency or "uncertainty." This perceived difficulty could provide a private cue that elicits use of the decline-test response. Forced test difficulty was initially titrated to produce a similar range of accuracy for each task. Still, each task may have elicited different subjective perceptions of difficulty across individual monkeys, perhaps as a result of different motivation to work and cognitive effort necessary to maintain accuracy. The control of metacognitive responding by subjective perception of difficulty is consistent with the task-by-task differences we observed in transfer. Individual differences in the use of the *decline-test* response are consistent with some human models of metacognition, which posit that just as cognition is not entirely accurate, metacognition is also subject to errors and individual differences in metacognitive sensitivity (Maniscalco and Lau, 2012; Nelson, 1996).

Consistent with individual variation in metacognitive accuracy, monkeys demonstrated substantial task-by-task variation in overall use of the *decline-test* response. Upon initial transfer 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. Some monkeys may have increased use of the *decline-test* response at transfer in part because use of the *decline-test* response was directly rewarded, a reinforcement contingency that differs from that used in some other paradigms (e.g., Smith et al., 2006; Couchman et al., 2010). Alternatively, novel tasks might be perceived as particularly difficult or effortful, with such a subjective,

private state increasing use of the decline-test response according to perceived difficulty even when objective difficulty, as measured by accuracy, is fairly constant.

10.4. Alternatives to monitoring of private information

Our transfer tasks were designed to dissociate control of metacognitive responding by private from control by public cues, but we cannot rule out the possibility that some public cues were shared across tasks. All primary psychophysical and memory tasks relied on experimenter-generated difficulty levels, which could cue reinforcement probability on a given trial. Such associative explanations have been proposed to account for apparently metacognitive behavior in nonhumans (Jozefowiez et al., 2009; Le Pelley, 2012). Our use of generalization tasks makes such an account unlikely, but not impossible. The overt cues that might signal low reinforcement probabilities are not consistent across memory and perceptual tasks. However, if monkeys develop "established response gradients" as they learn tasks, the *decline-test* response could be controlled by factors consistent across multiple tasks despite superficial task differences (Smith et al., 2008). We also cannot rule out the possibility that the appearance of test stimuli and the duration of the retention interval elicited similar self-generated public cues, which could control the metacognitive response. For instance, difficult trials, across domain, could elicit anxious behaviors associated with a low probability of food reward (Carruthers, 2008). In this way, the animal's external, publicly observable state could occasion use of the *decline-test* response. Such an account may beg the question of how the subject would become anxious without being directly sensitive to task difficulty.

11. Conclusions

The results presented here provide provisional evidence that rhesus monkeys use domain-general, private cues, such as "uncertainty," to monitor the status of cognitive processes and knowledge states, as has been proposed by Smith and colleagues (e.g., Smith et al., 2012). Much of the previous research on metacognition in nonhumans has focused on determining whether nonhumans manifest any metacognitive behavior, but less work has been devoted to identifying the mechanisms that underlie apparently metacognitive performance. Specifically, we do not know what cues or cognitive states control metacognitive responses. Because different paradigms have often been used in isolation, it has been unclear whether a domain-general private cue could account for metacognitive performance across diverse tasks. In comparing across previous work, it has been unclear whether cues controlling metacognitive behavior on perceptual tasks were the same cues controlling behavior on memory tasks. Even in instances when performance was likely controlled by a private cue (e.g., Hampton, 2001), it is unclear whether monkeys' behavior was controlled by a domain-general cue, like uncertainty, or a private cue specific to memory monitoring. One possibility generated by previous research was that monkeys attend to different cues depending on the nature of the primary cognitive task.

We are not the first to use transfer tests to assess the cues that control metacognitive responding. Monkeys generalized a concurrent "uncertain response" from discrimination learning to match-to-sample memory tests (Washburn et al., 2006), and also generalized retrospective "confidence" judgments across perceptual and serial order memory tests (Kornell et al., 2007). Use of retrospective metacognitive choices and transfer tests that demanded generalization across domains strengthen the claim that metacognitive responding was controlled by an internal state. Our test of generalization to a prospective metamemory judgment in

Experiment 6 further strengthens the hypothesis that metacognitive responding can be controlled in part by a private state, because the prospective metamemory judgment reduces the likelihood that metacognitive responding was controlled by behaviors like hesitation or vacillation that are typically elicited by difficult tests.

Here, we described a set of experiments that assessed the extent to which public or private cues control rhesus monkeys' metacognitive choices. Monkeys showed some evidence of generalization of the *decline-test* response across tasks. Because control of the decline-test response by task-specific cues would not predict rapid generalization, use of the *decline-test* response is not likely to be controlled exclusively by such publicly available, external cues in this case. However, when trained with any one task in isolation, task specific cues may well control metacognitive responding. Our data make a case for control of metacognitive responding by a cue that is both domain-general and private, such as "uncertainty", but the nature of this cue remains unclear. Future studies may focus on establishing refined characterizations of this domain-general private cue.

Statement on the welfare of animals: All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. All procedures performed in studies involving animals were in accordance with the ethical standards of the institution or practice at which the studies were conducted. This article does not contain any studies with human participants performed by any of the authors.

Conflict of interest

The authors declare that they have no conflict of interest.

Funding

Acknowledgements

References

Adams, A., Santi, A., 2011. Pigeons exhibit higher accuracy for chosen memory tests than for forced memory tests in duration matching-to-sample. Learn. Behav. 39, 1–11.

Basile, B.M., Hampton, R.R., 2014. Metacognition as discrimination: commentary on Smith et al. (2014). J. Comp. Psychol. 128, 135–137.

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. J. Exp. Psychol. Gen. 144, 85.

Brauer, J., Call, J., Tomasello, M., 2004. Visual perspective taking in dogs (*Canis familiaris*) in the presence of barriers. Appl. Anim. Behav. Sci. 88, 299–317, http://dx.doi.org/10.1016/j.applanim.2004.03.004.

Carruthers, P., 2008. Meta-cognition in animals: a skeptical look. Mind Lang. 23, 58–89, http://dx.doi.org/10.1111/j.1468-0017.2007.00329.x.

Couchman, J.J., Coutinho, M.V., Beran, M.J., Smith, J.D., 2010. Beyond stimulus cues and reinforcement signals: a new approach to animal metacognition. J. Comp. Psychol. 124, 356.

Flavell, J.H., 1979. Metacognition and cognitive monitoring: a new area of cognitive–developmental inquiry. Am. Psychol. 34, 906.

Foote, A.L., Crystal, J.D., 2007. Metacognition in the rat. Curr. Biol. 17, 551–555.

Fujita, K., 2009. Metamemory in tufted capuchin monkeys (*Cebus apella*). Anim. Cogn. 12, 575–585.

Goto, K., Watanabe, S., 2012. Large-billed crows (*Corvus macrorhynchos*) have retrospective but not prospective metamemory. Anim. Cogn. 15, 27–35.

Hampton, R.R., 2001. Rhesus monkeys know when they remember. Proc. Natl. Acad. Sci. U. S. A. 98, 5359–5362, http://dx.doi.org/10.1073/pnas.071600998.

Hampton, R.R., 2009a. Focusing the uncertainty about nonhuman metacognition. Comp. Cogn. Behav. Rev. 4, 56–57.

Hampton, R.R., 2009b. Multiple demonstrations of metacognition in nonhumans: converging evidence or multiple mechanisms? Comp. Cogn. Behav. Rev. 4, 17.

Inman, A., Shettleworth, S.J., 1999. Detecting metamemory in nonverbal subjects: a test with pigeons. J. Exp. Psychol. Anim. B 25, 389–395.

Iwasaki, S., Watanabe, S., Fujita, K., 2013. Do pigeons (*Columba livia*) seek information when they have insufficient knowledge? Anim. Cogn. 16, 211–221.

Jozefowiez, J., Staddon, J., Cerutti, D., 2009. Metacognition in animals: how do we know that they know. Comp. Cogn. Behav. Rev. 4, 29–39.

Kornell, N., Son, L.K., Terrace, H.S., 2007. Transfer of metacognitive skills and hint seeking in monkeys. Psychol. Sci. 18, 64–71, http://dx.doi.org/10.1111/j.1467-9280.2007.01850.x.

Kornell, N., 2013. Where is the meta in animal metacognition? J. Comp. Psychol. 128, 143–149, http://dx.doi.org/10.1037/a0033444.

Le Pelley, M.E., 2012. Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. J. Exp. Psychol. Learn. 38, 686–708, http://dx.doi.org/10.1037/a0026478.

Malassis, R., Gheusi, G., Fagot, J., 2015. Assessment of metacognitive monitoring and control in baboons (*Papio papio*). Anim. Cogn. 18, 1347–1362.

Maniscalco, B., Lau, H., 2012. A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. Conscious. Cogn. 21, 422–430.

Metcalfe, J., 2008. Evolution of metacognition. Handbook of metamemory and memory 29–46.

Morgan, G., Kornell, N., Kornblum, T., Terrace, H.S., 2014. Retrospective and prospective metacognitive judgments in rhesus macaques (*Macaca mulatta*). Anim. Cogn. 17, 249–257.

Nakamura, N., Watanabe, S., Betsuyaku, T., Fujita, K., 2011. Do birds (pigeons and bantams) know how confident they are of their perceptual decisions? Anim. Cogn. 14, 83–93.

Nelson, T.O., 1996. Consciousness and metacognition. Am. Psychol. 51, 102–116, http://dx.doi.org/10.1037//0003-066x.51.2.102.

Roberts, W.A., Feeney, M.C., McMillan, N., MacPherson, K., Musolino, E., Petter, M., 2009. Do pigeons (*Columba livia*) study for a test? J. Exp. Psychol. Anim. B 35, 129.

Roberts, W.A., McMillan, N., Musolino, E., Cole, M., 2012. Information seeking in animals: metacognition. Comp. Cogn. Behav. Rev. 7, 85–109.

Shettleworth, S.J., Sutton, J.E., 2003. Animal metacognition? It's all in the methods. Behav. Brain Sci. 26, 353.

Smith, J.D., Schull, J., Strote, J., McGee, K., Egnor, R., Erb, L., 1995. The uncertain response in the bottle-nosed dolphin (*Tursiops truncatus*). J. Exp. Psychol. Gen. 124, 391–408, http://dx.doi.org/10.1037//0096-3445.124.4.391.

Smith, J.D., Shields, W.E., Schull, J., Washburn, D.A., 1997. The uncertain response in humans and animals. Cognition 62, 75–97.

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. J. Exp. Psychol. Gen. 135, 282.

Smith, J.D., Beran, M.J., Couchman, J.J., Coutinho, M.V., 2008. The comparative study of metacognition: sharper paradigms, safer inferences. Psychon. Bull. Rev. 15, 679–691.

Smith, J.D., Redford, J.S., Beran, M.J., Washburn, D.A., 2010. Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. Anim. Cogn. 13, 93–101, http://dx.doi.org/10.1007/s10071-009-0249-5.

Smith, J.D., Couchman, J.J., Beran, M.J., 2012. The highs and lows of theoretical interpretation in animal-metacognition research. Philos. Trans. R. Soc. B 367, 1297–1309, http://dx.doi.org/10.1098/rstb.2011.0366.

Smith, J.D., Couchman, J.J., Beran, M.J., 2014. Animal metacognition: a tale of two comparative psychologies. J. Comp. Psychol. 128, 115.

Smith, J.D., 2009. The study of animal metacognition. Trends Cogn. Sci. 13, 389–396, http://dx.doi.org/10.1016/j.tics.2009.06.009.

Suda-King, C., Bania, A.E., Stromberg, E.E., Subiaul, F., 2013. Gorillas' use of the escape response in object choice memory tests. Anim. Cogn. 16, 65–84.

Suda-King, C., 2008. Do orangutans (*Pongo pygmaeus*) know when they do not remember? Anim. Cogn. 11, 21–42.

Templer, V.L., Hampton, R.R., 2012. Rhesus monkeys (*Macaca mulatta*) show robust evidence for memory awareness across multiple generalization tests. Anim. Cogn. 15, 409–419, http://dx.doi.org/10.1007/s10071-011-0468-4.

Washburn, D.A., Smith, J.D., Shields, W.E., 2006. Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. J. Exp. Psychol. Anim. B 32, 185–189, http://dx.doi.org/10.1037/0097-7403.32.2.185.

Washburn, D.A., Gulledge, J.P., Beran, M.J., Smith, J.D., 2010. With his memory magnetically erased, a monkey knows he is uncertain. Biol. Lett. 6, 160–162, http://dx.doi.org/10.1098/rsbl.2009.0737.

Wright, A.A., Katz, J.S., 2007. Generalization hypothesis of abstract-concept learning: learning strategies and related issues in *Macaca mulatta, Cebus apella,* and *Columba livia*. J. Comp. Psychol. 121, 387.