Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Anim Cogn. Author manuscript; available in PMC 2015 March 01.

Published in final edited form as:

Anim Cogn. 2014 March ; 17(2): 249–257. doi:10.1007/s10071-013-0657-4.

Retrospective and Prospective Metacognitive Judgments in Rhesus Macaques (*Macaca mulatta*)

Gin Morgan,

Department of Psychology, New Mexico State University, P.O. Box 30001, Las Cruces, New Mexico, 88003, Phone: 575-646-2502; Fax: 575-646-6212; gmorgan@nmsu.edu

Nate Kornell,

Department of Psychology, Williams College, 18 Hoxsey St., Williamstown, Massachusetts, 01267

Tamar Kornblum, and

Department of Psychology, Columbia University, 1190 Amsterdam Ave., New York, New York, 10027

Herbert S. Terrace

Department of Psychology, Columbia University, 1190 Amsterdam Ave., New York, New York, 10027

Abstract

A growing body of research suggests that some nonhuman animals are capable of making accurate metacognitive judgments. In previous studies, nonhuman animals have made either retrospective or prospective judgments (about how they did on a test or how they will do on a test, respectively). These two types of judgments are dissociable in humans. The current study tested the abilities of two rhesus macaque monkeys to make both retrospective and prospective judgments about their performance on the same memory task. Both monkeys had been trained previously to make retrospective confidence judgments. Both monkeys successfully demonstrated transfer of retrospective metacognitive judgments to the new memory task. Furthermore, both monkeys transferred their retrospective judgments to the prospective task (one, immediately, and one, following the elimination of a response bias). This study is the first to demonstrate both retrospective and prospective monitoring abilities in the same monkeys and on the same task, suggesting a greater level of flexibility in animals' metacognitive monitoring abilities than has been reported previously.

Keywords

metacognition; judgment; memory; monkeys

Correspondence to: Gin Morgan.

Ethical standards

The experiments described comply with the laws of the U.S.

Conflict of interest

The authors declare that they have no conflict of interest.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Metacognition refers to the ability to reflect on one's thoughts. Until recently, it was considered to be a key feature of conscious awareness and restricted to humans (Metcalfe and Shimamura 1994). However, recent studies suggest that some nonhuman animals (hereafter animals) also engage in metacognition—that is, they can correctly judge the accuracy of their perception and memory (Terrace and Metcalfe 2005).

Studies of metacognition in humans have tended to rely on verbal report—a clear challenge for measuring the subjective experiences of nonverbal animals. During the past fifteen years, however, an increasing number of studies have developed creative techniques for investigating metacognitive abilities in animals (e.g., Beran et al. 2006; Beran et al. 2009; Call and Carpenter 2001; Foote and Crystal 2007; Hampton 2001; Hampton et al. 2004; Inman and Shettleworth 1999; Kornell et al. 2007; Roberts et al. 2009; Shields et al. 2005; Smith et al. 1995; Smith et al. 1998; Washburn et al. 2006).

One of the key questions in the animal metacognition literature (and in animal cognition generally) is the degree to which animals' abilities resemble those of humans. Despite some similarities (e.g., Smith et al. 2003), humans demonstrate sophisticated abilities that have yet to be demonstrated in animals, such as the active selection of mnemonic strategies, and the ability to make complex, graded metacognitive judgments (e.g., Kornell 2009). One key aspect of human metacognition is that it is flexible: humans can make metacognitive judgments in almost any circumstance without the need for explicit training in each new context. The flexibility of animals' metacognitive abilities is therefore a key question, and recent research has shown that rhesus monkeys can indeed make metacognitive judgments across different tasks (Kornell et al. 2007) and that they seek different types of information necessary to complete a task (Beran and Smith 2011). Studies of flexibility are valuable in comparative metacognition because they allow for a broader comparison of animal abilities to those of humans, and because they may allow for more nuanced discriminations of metacognitive abilities between animal species (e.g., Beran and Smith 2011). The research presented here was designed to test flexibility by determining whether monkeys were able to make different metacognitive judgments on the same memory task (either preceding or following the test). An additional question we asked was whether performance would provide further evidence for monkeys' ability to transfer metacognitive skills between tasks.

The earliest research on metacognitive monitoring in animals used psychophysical paradigms to compare the use of an "uncertain" response by humans, monkeys, and a dolphin (Shields et al. 1997; Smith et al. 1995). The dolphin's task was to classify an auditory tone as low frequency or high frequency, and the monkeys' task was to classify a visual stimulus as either sparsely or densely pixelated. In both tasks, subjects could also opt to select an "uncertain" icon, which produced a reward (albeit one that was considerably delayed). Human participants were also trained on both the auditory and the visual tasks. A comparison of the performance of the three species revealed some striking similarities. Each species used the "uncertain" response appropriately—that is, more often on trials near their discriminatory threshold.

Similar use of the "uncertain" response does not, of course, imply that the use of that response by different species was a manifestation of the same underlying mechanism. The

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

uncertainty response may have resulted from a learned association between particular stimuli and behavioral responses that maximized reward. In order to counter behavioral explanations of uncertainty responses, Smith et al. (2008) recommended several modifications of the metacognitive paradigms used with animals. The most important were: 1) to demonstrate immediate transfer of metacognitive skills to new, qualitatively different, tasks, 2) the use of abstract cognitive domains, and 3) the use of cognitive judgments (e.g., a reliance on internal, rather than exteroceptive, cues). Similarly, Terrace and Son (2009) stressed the importance of not allowing escape, or "uncertain," responses to be made in the presence of the stimuli on which animals are trained, as otherwise, it would not be possible to rule out stimulus control by the training stimuli. Terrace and Son also argued that escape or "uncertain" responses should not be allowed to increase the task's reinforcement rate beyond what would ordinarily be the maximum reward (for instance, by replacing a difficult trial with an easy one that would be more likely to yield a reward). This problem can be avoided if using the escape or uncertain response does not replace a trial, but instead, simply moves to the next trial (e.g., Smith et al. 2006; Smith et al. 2010), or if the uncertain response, when used indiscriminately, does not result in maximization of reward. This problem can also be avoided when increases in reinforcement rate are accompanied by penalties for overusing the uncertain response (in the current experiments, a delay was imposed to prevent such bias).

Research on human metacognition employs a variety of different types of judgments, but one of the basic divisions is whether judgments are about the past or the future. Retrospective judgments are made about events that were previously experienced, e.g., confidence judgments ("I'm sure that I was right"), remember/know judgments ("I remember the moment at which I learned this information"), and source monitoring judgments ("I know because I heard it on the radio"). By contrast, prospective judgments refer to judgments of confidence about future events. Examples include feeling of knowing judgments ("I'm sure that I would recognize the answer") and tip-of-the-tongue judgments ("I know the answer but cannot produce it").

Evidence of both prospective and retrospective metacognition has been obtained from animals. For example, Hampton (2001) and Fujita (2009) investigated prospective metamemory judgments using a delayed matching-to-sample paradigm. Following the presentation of the sample, but before test, monkeys were allowed, on some trials, to choose whether to proceed to test or to opt out of taking the test. Correct completion of the test resulted in a more desirable food reward whereas opting out of the test resulted in a less desirable but guaranteed food reward. In both tasks, monkeys performed better on chosen trials (those on which they had chosen to proceed to test) than on forced trials. The monkeys also chose to opt out more often on trials when no sample was presented. These findings suggest that the monkeys opted out of trials on which they did not know the correct response and that monkeys are therefore capable of making appropriate prospective metacognitive judgments.

Kornell et al. (2007) tested retrospective metamemory judgments using a paradigm that measured monkeys' confidence by asking them to "bet" on the responses they had just made on a memory task or a perceptual task (see Shields et al., 2005, for similar findings using a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

perceptual task). Kornell et al. first trained two monkeys on perceptual tasks (discriminating line length, numerosity, and circle size). After making a response, but before feedback was given, the monkeys were required to select one of two icons: low or high risk. Selection of the low risk icon resulted in a small, guaranteed reward regardless of performance on that trial. Selection of the high risk icon resulted in a comparatively large gain if the monkey had completed the trial correctly, or an equally large loss if the monkey was incorrect on that trial. Rewards were presented or removed by using a computerized token economy, and a food reward was delivered once the balance of the token bank reached a certain threshold. After mastering the three perceptual tasks, the animals were trained on a memory task. On all four tasks, monkeys successfully maximized their reward by selecting low risk on trials that they had completed incorrectly, and high risk on trials that they had completed correctly. Importantly, monkeys showed immediate transfer of metacognitive skills between tasks, indicating that their metacognitive skills were flexible in the sense that they were not specific to a particular task.

Although prospective and retrospective metamemory tasks have been used successfully in animals, they have not been examined in the same animals and on the same tasks. There are reasons to expect that the distinction between prospective and retrospective judgments is important. A recent review of neuropsychological evidence from humans suggests that the medial prefrontal cortex (PFC) is involved in prospective judgments, whereas the lateral PFC is involved in retrospective judgments (see Fleming and Dolan 2012 for a review; also see Nelson, 1996). Functional and neurological dissociations have been found in populations such as Korsakoff patients (who have frontal lobe damage) and high-altitude climbers (in whom hypoxia-induced cognitive impairments have been found; see Nelson et al. 1990, for a review). In one study, Korsakoff patients demonstrated impairments in feeling-of-knowing accuracy (a prospective judgment) but no impairments in confidence judgment accuracy (a retrospective judgment; Shimamura and Squire 1986, 1988). High-altitude climbers tested with the same materials showed a similar dissociation between feeling-of-knowing judgments and confidence judgments (Nelson, 1996; Nelson et al. 1990). These studies suggest that when human frontal lobe function is compromised, prospective judgments suffer but retrospective judgments remain intact. It is worth noting that monkeys have comparatively less frontal lobe volume than do humans (Semendeferi et al. 2002). An examination of prospective and retrospective judgments in animals has the potential to further identify similarities, or differences, between human and animal metacognition.

Two experiments were designed to test the flexibility of monkeys' metacognitive monitoring and to examine monkeys' retrospective and prospective metamemory judgments. In both experiments, risk choices were used to measure the monkeys' certainty about their memories. Experiment 1 tested monkeys' use of retrospective metamemory: the monkeys made risk choices *after* they had completed a memory test. Experiment 2 tested monkeys' use of prospective metamemory: the risk choices were presented after the sample presentation (i.e., after learning) but *before* the memory test. In both cases, the metamemory judgments were not made in the presence of either learning or test stimuli. Thus, subjects' responses could not be based on external stimuli, and were therefore more likely to be based on the content of their internal memories (Kornell et al. 2007; Smith et al. 2008; Terrace and

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Son, 2009). In both experiments, we expected that monkeys would tend to make low risk judgments when they were incorrect and high risk judgments when they were correct. Alternatively, if frontal lobe capacity were a limiting factor in monkeys' metacognitive abilities, we might expect accurate metacognitive performance on the retrospective task (Experiment 1) but not on the prospective task (Experiment 2).

Experiment 1: Transfer of retrospective metacognitive monitoring skills from previous tasks

Experiment 1 was designed to test whether two monkeys, previously trained to make retrospective risk choices, would transfer this ability to a new memory task.

Method

Subjects and apparatus—Two male rhesus monkeys (*Macaca mulatta*), Ebbinghaus and Lashley, were tested. Ebbinghaus and Lashley had extensive experience with metacognitive tasks (see Kornell et al. 2007; Son and Kornell 2005). They had previously made retrospective risk choices following test on a series of psychophysical tasks and on a memory task. Neither monkey was tested concurrently on additional metacognitive tasks during the metacognitive testing phase of the current task.

Subjects were housed individually in a rhesus colony room and were transported to a testing room in individual transport boxes. Subjects were monitored by a full-time veterinary staff, and all housing, transport, and testing procedures were conducted in accordance with IACUC guidelines. Subjects were fed Purina monkey chow (Ralston Purina, Richmond, IN) and fresh fruit following the completion of their testing. Water was available *ad libitum* in subjects' home cages.

Subjects completed one 60 trial session per day that lasted approximately 15 minutes. Each experimental chamber was housed in an isolated sound-attenuated booth and contained a 3M MicroTouch touch-sensitive video monitor (3M, St. Paul, MN) and a Gerbrands pellet dispenser (Med Associates, Inc., Georgia, VT). During the task, monkeys received 190-mg Noyes banana-flavored pellets (P. J. Noyes, Lancaster, NH) as food rewards.

Stimuli—The stimuli were pictures of various manmade and natural objects that were selected at random from a pool of approximately 2,000 pictures. While monkeys saw each picture multiple times over the weeks of training and testing, the same picture was never repeated on the same day and it was highly unlikely that the stimuli on a given trial were ever repeated in the same combination.

Procedure—Before training began, the subjects had been trained on a computerized token economy in which they had to earn tokens in order to gain food rewards. A two-dimensional bank on the lower-right corner of the screen contained six tokens at the outset of each session. Tokens could be gained or lost based on the subject's responses (see below). When the number of tokens reached or exceeded eight, the monkey received two 190-mg banana-flavored food pellets and the level of tokens was reset to six.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Monkeys were tested on a working memory task using a modified delayed matching-to-sample design (Figure 1a). During the sample presentation, monkeys were presented with multiple images on the screen simultaneously (2 samples/trial for Lashley; 3 samples/trial for Ebbinghaus) for 2500 msec. At the end of the sample presentation (during the last 2000 msec for Lashley; 1250 msec for Ebbinghaus), a moving border appeared around one of the samples, indicating that it should be chosen at test. Test immediately followed the sample presentation. The test consisted of all the samples presented on that trial in addition to multiple distractors (3 for Lashley; 5 for Ebbinghaus). These different task parameters (number of samples, number of distractors, and sample presentation times) were manipulated for each monkey during a training phase in order to equate task difficulty between monkeys. This resulted in session accuracies between 40% and 70% on the matching-to-sample task. The same criteria were used later during the metacognitive test phase.

Monkeys were first trained on the memory task without making confidence judgments (approximately 300 sessions for Ebbinghaus and 540 sessions for Lashley). During this training phase, feedback regarding the accuracy of a trial was given immediately following test. When a trial was completed correctly, two tokens were delivered into the monkey's token bank, and when the number of tokens reached eight, the monkey received a food reward. When a trial was completed incorrectly, two tokens were removed from the monkey's token bank. The goal of this training period was to ensure that both monkeys learned the memory task before the introduction of the confidence judgment.

After the initial training phase on the memory task, a confidence judgment was introduced such that monkeys made either a low risk or high risk "bet" on each trial. After test, but before feedback regarding that trial's accuracy, the low and high risk icons appeared on the screen. A low risk response was rewarded with one token regardless of the accuracy of the monkey's response on that trial. When a high risk response was made, the monkey received three tokens if his response on that trial was correct, but lost three tokens if his response was incorrect. Thus, reward was maximized if the subject chose high risk after a correctly completed trial and low risk after an incorrectly completed trial. As before, the default value of the token bank was six, and when the number of tokens in the bank reached eight, two food pellets were delivered and the bank balance was reset to six tokens.

The task used a bias reduction procedure in which the risk icon that was selected less frequently during that session was immediately available. The risk icon to which the subject responded more frequently was presented after a delay. The length of the delay was modified on every trial to reflect the extent of the bias. As the bias increased, the delay increased as determined by the following formula:

$$\text{delay} = [(\text{previous bias value})(.97) +/- 1]/2$$

such that the bias was increased or decreased on each trial by .5 sec. The beginning delay value for each session was typically set as the ending delay value from the previous session. For the first five sessions following the introduction of the risk icons, both monkeys showed slight high risk biases. The maximum delays for Lashley and Ebbinghaus during this period were 3.37 sec and 5.98 sec (both biased toward high risk), respectively. Ebbinghaus'

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

average bias at the end of each session was 2.17 sec toward high risk. Lashley's average end-session delay was 0.24 sec (also toward high risk). While this delay encouraged selection of the less-chosen icon, it did not provide any information as to which icon was appropriate to select on a given trial.

Results

Data from the first five sessions following the introduction of the risk icons were analyzed to determine whether monkeys successfully transferred metacognitive skills from previously completed tasks to the current memory task. To determine the metacognitive accuracy of monkeys' confidence judgments, we calculated phi correlations between accuracy and risk choice. In the first five sessions following the introduction of the confidence judgment, both Ebbinghaus and Lashley showed significant correlations between accuracy and risk choice (Ebbinghaus, phi = .36, $p < .001$; Lashley, phi = .18, $p = .001$), as both monkeys tended to select low risk following incorrectly completed trials, and high risk following correctly completed trials. Ebbinghaus' accuracy on trials on which he selected low risk was 42%, while his accuracy on trials on which he selected high risk was 77%. Lashley's accuracy on low risk trials was 60%, and his accuracy on high risk trials was 77% (Figure 2). In both cases, session accuracy when low risk was selected was significantly higher than chance performance (Ebbinghaus, $t(4) = 5.72$, $p = .005$; Lashley, $t(4) = 7.28$, $p = .002$). This suggests that monkeys were not likely deciding, in advance, to select low risk on a given trial. Instead, the monkeys appeared to follow the more advantageous strategy to maximize reward: attempting (generally) all trials, and basing their subsequent risk choice on perceived performance.

One possible explanation for the differential use of the risk choices is that subjects simply chose low risk when they had responded slowly on that trial—that is, that reaction time (RT) served as a discriminative stimulus. Because both monkeys showed a significant correlation between RT and risk choice (Pearson correlations, Ebbinghaus, r = −.13, $p = .03$, Lashley, r = −.20, $p = .001$), we calculated partial correlations between RT at test, accuracy, and risk choice. After controlling for RT, a significant correlation remained between accuracy and risk choice for both monkeys (Ebbinghaus, r = .34, $p < .001$; Lashley, r = .15, $p = .01$), indicating that RT was not the sole cue for monkeys' risk choices.

These results demonstrate that monkeys are adept at monitoring their performance and responding appropriately on a retrospective metamemory task. Additionally, because monkeys immediately responded appropriately following the introduction of the confidence judgment—that is, their metacognitive judgment transferred from their previous training to this task—these results provide additional evidence that monkeys can learn flexible metacognitive monitoring skills that are not task-specific.

Experiment 2: Testing prospective metacognitive monitoring skills

In Experiment 1, Ebbinghaus and Lashley made appropriate retrospective confidence judgments on a memory task. Experiment 2 was designed to test whether, using the same memory task, these monkeys would make appropriate prospective confidence judgments. The task used in Experiment 1 was altered for Experiment 2 such that the risk judgment was

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

made before test (rather than following test; see Figure 1b). Token feedback was provided after the memory choice was made, rather than immediately following the risk choice.

Method

Subjects and apparatus—The monkeys and apparatus in Experiment 2 were the same in Experiment 1. Neither monkey was tested concurrently on an additional metacognitive task, with the exception of Ebbinghaus' testing on both retrospective and prospective judgments, described below. Neither monkey had prior experience making prospective risk choices.

Procedure—After the introduction of the confidence judgment, both monkeys continued training on the retrospective task from Experiment 1 (approximately 60 sessions for Ebbinghaus and 160 sessions for Lashley), and adjustments to the delay between sample offset and test onset were made. Because confidence judgments on the prospective task occurred after the sample presentation and before test, the prospective task involved a longer delay between sample offset and test onset than did the retrospective task. Thus, to avoid decrements in performance upon transfer to the prospective judgment, a delay was gradually introduced to the retrospective task that ultimately approximated the delay they would experience during the prospective task (2.25 sec).

Experiment 2 began after accuracy on the retrospective task with the longer delay stabilized around 55% (task accuracy for the last 10 sessions of the retrospective task was within 1 standard deviation of task accuracy across all training on the retrospective task). Two minor methodological changes were made to the sample presentation in order to achieve similar task accuracy levels between the two monkeys. First, although the sample presentation lasted 2500 msec, as in Experiment 1, the moving border was now present for 1250–1500 msec for Ebbinghaus and 1600 msec for Lashley. Second, although Ebbinghaus' initial testing in Experiment 2 involved 3 stimuli during the sample presentation, this value was changed such that, in subsequent testing (after the introduction of the bias reduction procedure), both Ebbinghaus and Lashley viewed 2 stimuli during the sample presentation.

When the prospective risk judgment was initially introduced, the bias reduction procedure used in Experiment 1 was not used, in order to avoid overly long delays between sample presentation and test. In the retrospective task in Experiment 1, there was no delay between sample presentation and test. In the prospective task in Experiment 2, there was necessarily a short delay between sample presentation and test (so that the monkeys could make a risk judgment before test). Incorporating the bias reduction procedure between sample presentation and test could create a delay of over 10 seconds before test, thereby increasing task difficulty. Thus, to achieve our original goal of similar task difficulties across Experiments 1 and 2, we began the prospective judgment phase of Experiment 2 without the bias reduction procedure.

During the first ten sessions following the introduction of the prospective judgment, only Ebbinghaus utilized the high risk option, albeit at a lower rate than on the retrospective task. Lashley did not make a single high risk response during the first ten sessions of testing. Ebbinghaus began testing on Experiment 2 before Lashley, and after the initial ten sessions requiring prospective judgments, two steps were taken to reduce Ebbinghaus' response bias.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

First, Ebbinghaus completed approximately 45 days of training in which he completed one 30-trial session of the retrospective task and one 30-trial session of the prospective task. Because this training was not successful in reducing his response bias on the prospective task, the bias reduction procedure used in Experiment 1 was introduced. Because both monkeys were biased toward making low risk responses during their initial testing on the prospective task, each session began with a pre-specified delay before the low risk icon became available (Ebbinghaus, 2 sec; Lashley, 1 sec). After the first trial, the delay was updated from trial to trial based on each monkey's response pattern. The introduction of the bias reduction procedure resulted in an immediate improvement in Ebbinghaus' performance on the prospective task, although he continued to be tested on a 30-trial retrospective session and a 30-trial prospective session each day. After Lashley completed his first ten sessions with the prospective judgment without the bias reduction procedure, the bias reduction procedure was introduced to his task as well (although he completed only the prospective task, rather than both the retrospective and prospective tasks simultaneously). Contrary to the high risk bias both monkeys showed on the retrospective task, both showed a low risk bias on the prospective task. During the first five sessions, Lashley had a maximum delay of 10.76 sec compared to Ebbinghaus' maximum delay of 5.29 sec. Comparing risk biases at the end of each session, however, results for each monkey were similar; Ebbinghaus' average low risk bias was 2.02 sec, while Lashley's average low risk bias was 1.45 sec. Results from both the initial five-session transfer period and the first five sessions after the introduction of the bias reduction procedure are presented below.

Results

To determine the metacognitive accuracy of monkeys' confidence judgments, we calculated phi correlations between accuracy and risk choice. During the first five sessions following the introduction of the confidence judgment, Ebbinghaus' correlation between accuracy and risk choice was significant (phi = .12, $p = .04$). A correlation could not be computed for Lashley because he never chose high risk. Lashley's extreme low risk bias during the first five sessions was likely due to motivational issues rather than task difficulty; his accuracy on the first five sessions following the introduction of the confidence judgment (51%) was similar to his accuracy on the final five sessions of the retrospective task (45%). Lashley's phi correlation during the final five sessions of the retrospective task was .30 ($p < .001$), indicating an ability to make metacognitively accurate judgments at a similar level of task accuracy. As on the retrospective task, session accuracy for low risk responses was higher than chance performance (Ebbinghaus, $M = 33\%$, $t(4) = 5.98$, $p = .004$; Lashley, $t(4) = 6.18$, $p = .003$), suggesting that both monkeys were still attempting the task, even on trials on which they selected low risk.

During the first five sessions after the bias reduction procedure was introduced, both Ebbinghaus' and Lashley's performance was characterized by significant correlations between accuracy and risk choice (Ebbinghaus, phi = .26, $p = .002$; Lashley, phi = .21, $p < .001$). Both monkeys tended to select low risk before making an incorrect response, and high risk before making a correct response. Ebbinghaus' accuracy on trials on which he selected low risk was 26%, while his accuracy on trials on which he selected high risk was 51%. Lashley's accuracy on low risk trials was 32%, and his accuracy on high risk trials was 53%

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

(Figure 3). Ebbinghaus' session accuracy for low risk trials was significantly above chance ($t(4) = 3.15$, $p = .04$), but Lashley's session accuracy for low risk trials was only marginally above chance ($t(4) = 2.29$, $p = .08$). A possible cause of the lower task accuracy (compared to task accuracy on the retrospective task) was the variable delay between sample presentation and test introduced by the bias-reduction procedure. In contrast to the retrospective task, reaction time was not a plausible cue for selecting low or high risk, since the risk choice was made before test.

In sum, these results demonstrate that one monkey successfully transferred metacognitive skills from a retrospective task to a prospective task immediately. The other monkey showed transfer as soon as his low risk bias was reduced. Thus, both monkeys showed strong evidence of prospective metamemory abilities.

General discussion

The results presented in Experiments 1 and 2 add to a growing body of research that has demonstrated metacognitive abilities in monkeys. In Experiment 1, which involved a retrospective confidence judgment, both monkeys showed immediate transfer of metacognitive skills to a new memory task. In Experiment 2, which involved a prospective metamemory judgment, both monkeys showed transfer to the new task as soon as it was possible to measure their metacognitive abilities (i.e., when their bias was not absolute).

These experiments are the first to demonstrate both retrospective and prospective judgments using the same memory task and the same monkeys. Because retrospective and prospective judgments are dissociable in humans, it follows that success on one type of task does not necessarily predict success on the other. Nevertheless, both monkeys were able to make both types of judgments. Further research could clarify whether dissociation between retrospective and prospective metamemory does occur among monkeys and other animals, or if monkeys tend to demonstrate a greater degree of metacognitive competence than their frontal lobe volume might otherwise suggest. Taken together, these results demonstrate a new degree of flexibility in animals' metacognitive abilities (see Smith et al. 2010; Washburn et al. 2006). The monkeys transferred their metacognitive monitoring skills to a new memory task, thereby replicating transfer abilities described by Kornell et al. (2007). Moreover, they were able to switch between different *types* of metacognitive monitoring skills.

The two experiments of this study incorporated several important design features suggested by Smith et al. (2008) and by Terrace and Son (2009). First, Experiments 1 and 2 tested transfer of metacognitive abilities between tasks. Second, the tasks involved judgments of memory tasks—in the case of Experiment 1, a judgment of how well the monkey remembered the probe stimulus on that trial, and in the case of Experiment 2, how likely the monkey was to remember the probe stimulus at test; thus, judgments were not made in the presence of exteroceptive stimuli. Third, every trial during a session consisted of the same numbers of sample and distractor stimuli, so the number of samples or test stimuli could not serve as a cue on which monkeys could base their risk judgments. Last, the use of the low

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

risk response did not result in an above-maximal reinforcement rate because a low risk response did not lead to the replacement of that trial with a potentially easier trial.

This study contributes to a growing understanding of animal metacognitive abilities, but much remains to be investigated. Are animals capable of making complex, graded metacognitive judgments? To what extent do animals guide their own learning? Above all, how do animals behave metacognitively, and are their metacognitive judgments entirely implicit? These are just a few areas of exploration that can shed light on the phylogenetic origins of metacognition. The flexibility of monkeys' metacognitive abilities described in these studies yields further evidence of comparable metacognitive abilities in animals and humans.

Acknowledgements

The project described was supported by award number 5R01MN081153NIMH from the National Institutes of Mental Health and award number JSMF 220020088 from the James S. McDonnell foundation to HST. We gratefully acknowledge the assistance of Niko Reyes and the graduate students and research assistants of the Primate Cognition Lab.

References

Beran MG, Smith JD. Information seeking by rhesus monkeys (Macaca mulatta) and capuchin monkeys (Cebus apella). Cogn. 2011; 120:90–105.

Beran MJ, Smith JD, Redford JS, Washburn DA. Rhesus macaques (Macaca mulatta) monitor uncertainty during numerosity judgments. J Exp Psychol Anim B. 2006; 32(2):111–119.

Beran MJ, Smith JD, Coutinho MVC, Couchman JJ, Boomer J. The psychological organization of "uncertainty" responses and "middle" responses: A dissociation in capuchin monkeys (Cebus apella). J Exp Psychol Anim B. 2009; 35(3):371–381.

Call J, Carpenter M. Do apes and children know what they have seen? Anim Cogn. 2001; 4:207–220.

Fleming SM, Dolan RJ. The neural basis of metacognitive ability. Phil Trans R Soc B. 2012; 367:1338–1349. [PubMed: 22492751]

Foote AL, Crystal JD. Metacognition in the rat. Curr Biol. 2007; 17:1–5. [PubMed: 17208179]

Fujita K. Metamemory in tufted capuchin monkeys (Cebus apella). Anim Cogn. 2009; 12:575–585. [PubMed: 19242741]

Hampton RR. Rhesus monkeys know when they remember. P Natl Acad Sci USA. 2001; 98(9):5359–5362.

Hampton RR, Zivin A, Murray EA. Rhesus monkeys (Macaca mulatta) discriminate between knowing and not knowing and collect information as needed before acting. Anim Cogn. 2004; 7:239–246. [PubMed: 15105996]

Inman A, Shettleworth SJ. Detecting metamemory in nonverbal subjects: a test with pigeons. J Exp Psychol Anim B. 1999; 25(3):389–395.

Kornell N. Metacognition in humans and animals. Curr Dir Psychol Sci. 2009; 18(1):11–15.

Kornell N, Son LK, Terrace HS. Transfer of metacognitive skills and hint seeking in monkeys. Psychol Sci. 2007; 18(1):64–71. [PubMed: 17362380]

Metcalfe, J.; Shimamura, AP., editors. Metacognition: Knowing about knowing. Cambridge, MA: Bradford Books; 1994.

Nelson TO. Consciousness and Metacognition. Am Psychol. 1996; 51(2):102–116.

Nelson TO, Dunlosky J, White DM, Steinberg J, Townes BD, Anderson D. Cognition and metacognition at extreme altitude on Mount Everest. J Exp Psychol Gen. 1990; 119:367–374. [PubMed: 2148573]

Roberts WA, Feeney MC, McMillan N, MacPherson K, Musolino E, Petter M. Do pigeons (Columba livia) study for a test? J Exp Psychol Anim B. 2009; 35(2):129–142.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Semedenferi K, Lu A, Schenker N, Damasio H. Humans and great apes share a large frontal cortex. Nat Neurosci. 2002; 5(3):272–276. [PubMed: 11850633]

Shields WE, Smith JD, Guttmannova K, Washburn DA. Confidence judgments by humans and rhesus monkeys. J Gen Psychol. 2005; 132(2):165–186. [PubMed: 15871299]

Shields WE, Smith JD, Washburn DA. Uncertain responses by humans and rhesus monkeys (Macaca mulatta) in a psychophysical same-different task. J Exp Psychol Gen. 1997; 126(2):147–164. [PubMed: 9163934]

Shimamura AP, Squire LR. Memory and metamemory: A study of the feeling-of-knowing phenomenon in amnesic patients. J Exp Psychol Learn. 1986; 12(3):452–460.

Shimamura AP, Squire LR. Long-term memory in amnesia: Cued recall, recognition memory, and confidence ratings. J Exp Psychol Learn. 1988; 14(4):763–770.

Smith JD, Beran MJ, Couchman JJ, Coutinho MVC. The comparative study of metacognition: Sharper paradigms, safer inferences. Psychon B Rev. 2008; 15(4):679–691.

Smith JD, Beran MJ, Redford JS, Washburn DA. Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. J Exp Psychol Gen. 2006; 135(2):282–297. [PubMed: 16719654]

Smith JD, Redford JS, Beran MJ, Washburn DA. Rhesus monkeys (Macaca mulatta) adaptively monitor uncertainty while multi-tasking. Anim Cogn. 2010; 13:93–101. [PubMed: 19526256]

Smith JD, Schull J, Strote J, McGee K, Egnor R, Erb L. The uncertain response in the bottlenosed dolphin (Tursiops truncatus). J Exp Psychol Gen. 1995; 124(4):391–408. [PubMed: 8530911]

Smith JD, Shields WE, Allendorfer KR, Washburn DA. Memory monitoring by animals and humans. J Exp Psychol Gen. 1998; 127(3):227–250. [PubMed: 9742715]

Smith JD, Shields WE, Washburn DA. The comparative psychology of uncertainty monitoring and metacognition. Behav Brain Sci. 2003; 26:317–373. [PubMed: 14968691]

Son, LK.; Kornell, N. Meta-confidence judgments in rhesus macaques: Explicit versus implicit mechanisms. In: Terrace, HS.; Metcalfe, J., editors. The missing link in cognition: Origins of self-reflective consciousness. New York: Oxford University Press; 2005. p. 296-320.

Terrace, HS.; Metcalfe, J., editors. The Missing Link in Cognition: Origins of Self-Reflective Consciousness. New York: Oxford University Press; 2005.

Terrace HS, Son L. Comparative Metacognition. Curr Opin Neurobiol. 2009; 19:67–74. [PubMed: 19553102]

Washburn DA, Smith JD, Shields WE. Rhesus monkeys (Macaca mulatta) immediately generalize the uncertain response. J Exp Psychol Anim B. 2006; 32(2):185–189.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

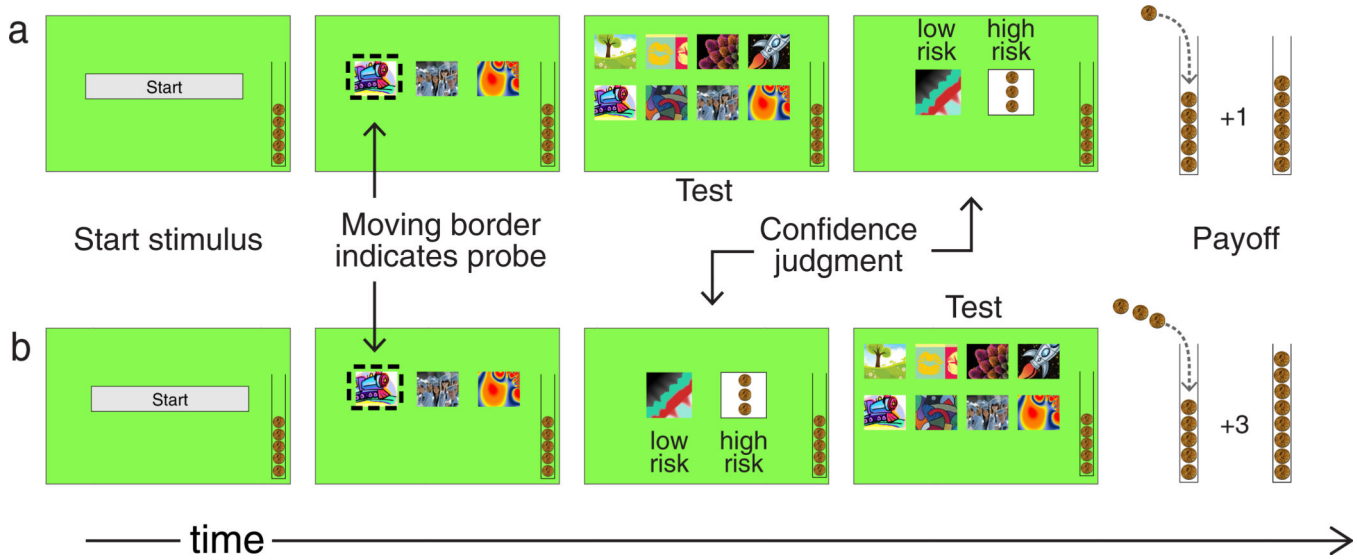


Fig. 1.

Schematic of the delayed matching-to-sample procedure used a) in the retrospective task in Experiment 1 and b) in the prospective task in Experiment 2.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

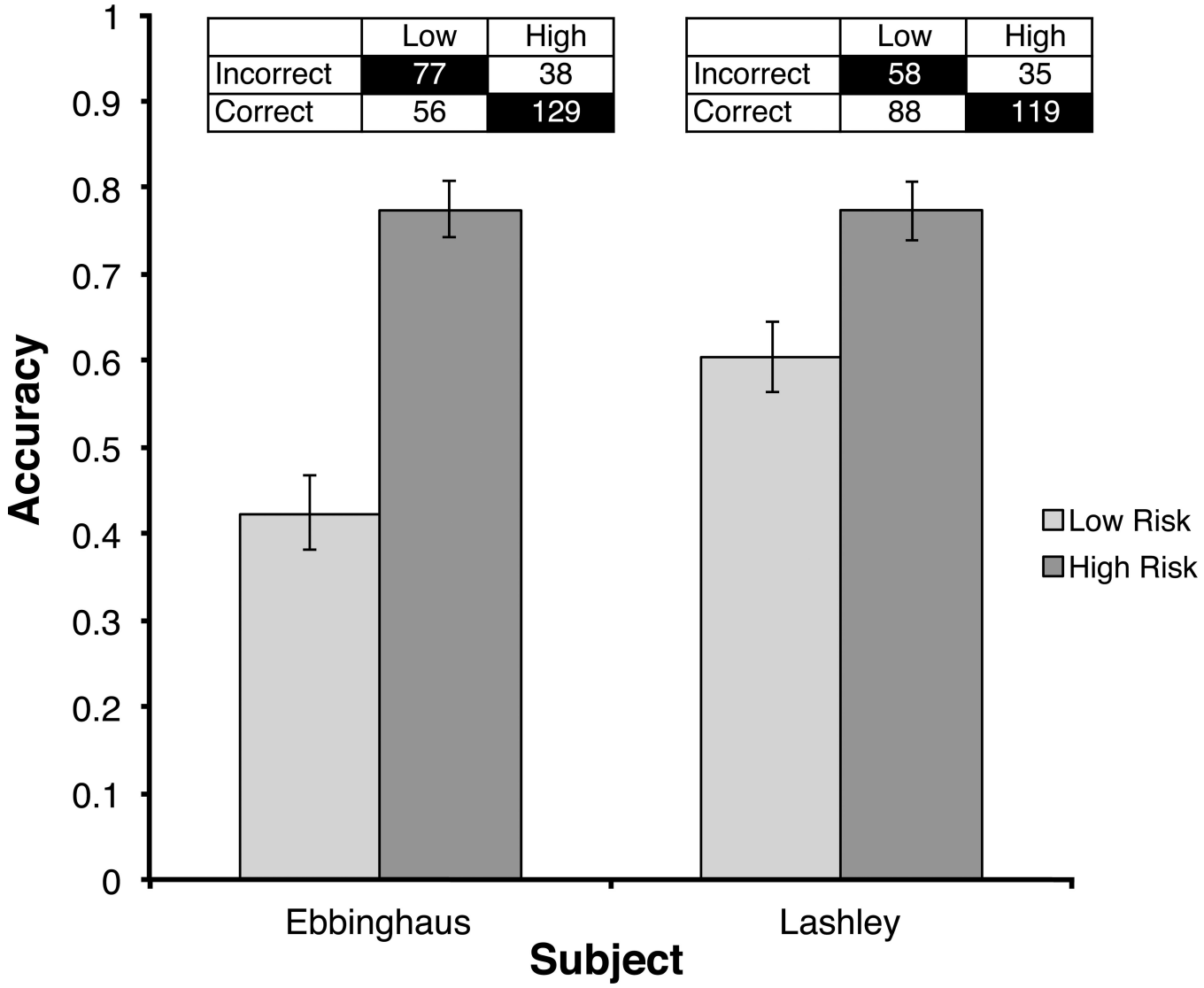
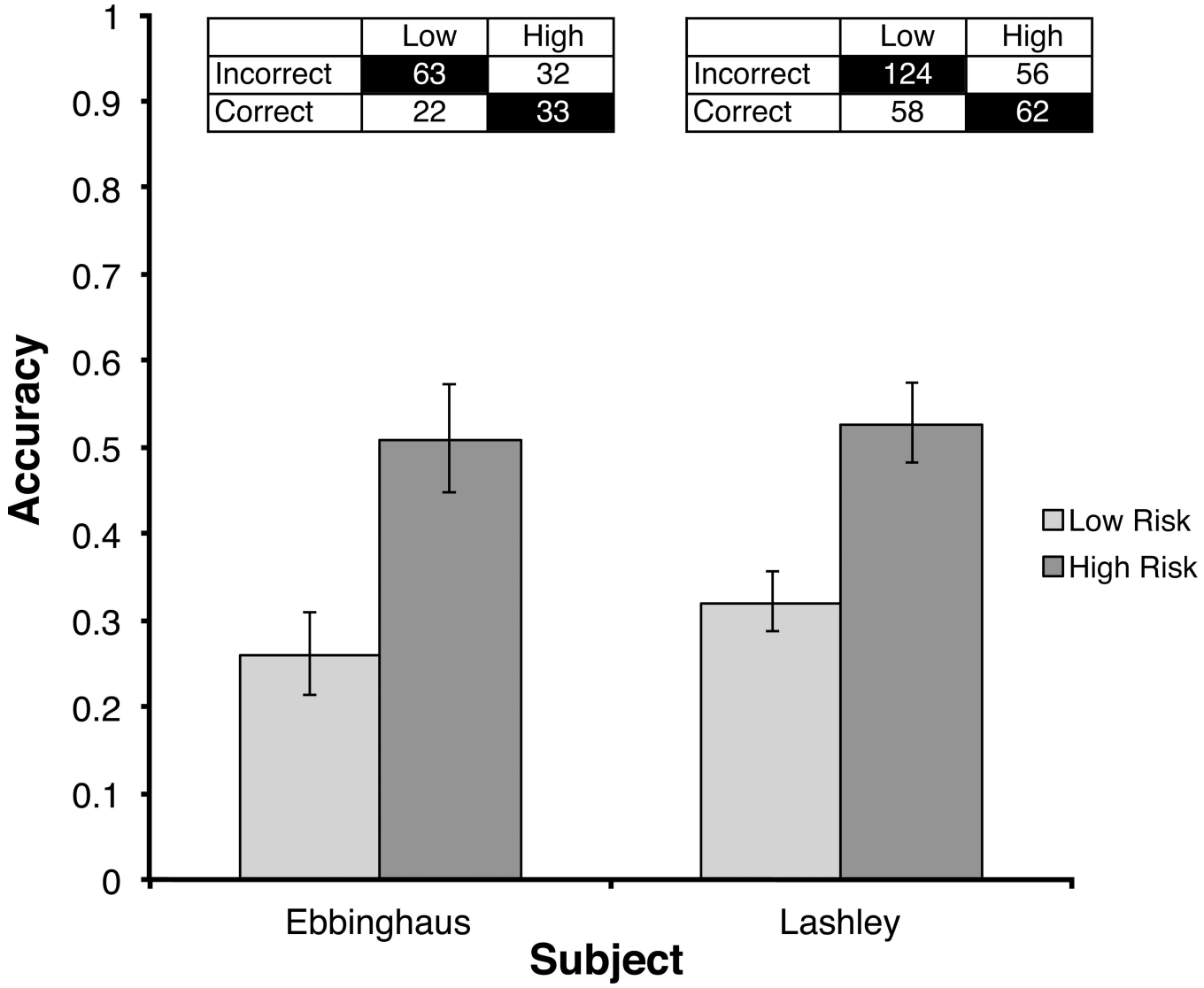


Fig. 2.

Average accuracy when low and high risk retrospective judgments were made. The number of low and high risk bets following incorrect and correct responses are listed for each monkey. Cells with a black background indicate metacognitively appropriate responses. Data are from the first 5 days following the introduction of the retrospective risk judgment.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript



	Low	High
Incorrect	63	32
Correct	22	33

	Low	High
Incorrect	124	56
Correct	58	62

Fig. 3.

Average accuracy when low and high risk prospective judgments were made. The number of low and high risk bets following incorrect and correct responses are listed for each monkey. Cells with a black background indicate metacognitively appropriate responses. Data are from the first 5 days following the introduction of the bias elimination procedure.