



Exploring Potential Mechanisms Underlying the Lack of Uncertainty Monitoring in Capuchin Monkeys

Bonnie M. Perdue¹, Barbara A. Church², J. David Smith², & Michael J. Beran³

¹Department of Psychology, Agnes Scott College, USA

²Department of Psychology, University at Buffalo, The State University of New York, USA

³Language Research Center and Department of Psychology, Georgia State University, USA

In a widely used animal-metacognition paradigm, monkeys are positively reinforced with food for correct classifications of stimuli as sparse or dense and punished with timeouts for incorrect responses, but they also have access to an *uncertainty* response that moves them to the next trial without either of these forms of feedback. Rhesus monkeys use this uncertainty response most often for trials on which they are at greatest risk for making an error, suggesting that they are monitoring their ability to make these classifications. Capuchin monkeys do not succeed to the same degree on these tasks - conceivably as a result of differential contingencies in place in all existing studies between the sparse/dense responses (food delivery or timeout) and the uncertainty response (avoidance of a timeout but also no chance for food reward). Here, we used a novel variation of this task in which the outcomes of the three response classes (sparse, dense, uncertain) were functionally equivalent. All responses simply determined the delay interval before presentation of a second task (matching-to-sample), and that task yielded potential food rewards. Overall, capuchin monkeys used the dense and sparse responses appropriately, including some animals that had no prior experience in performing this classification task. However, none used the uncertainty response appropriately even when it was placed on the same contingency plane as the dense and sparse responses. This suggests that the failure of capuchin monkeys to use an uncertainty response is not the result of that response producing a qualitatively different outcome compared to the dense and sparse responses.

Humans are metacognitive - often expressed as an explicit, declarative awareness of their states of learning and knowing. Metacognition is a late developmental achievement (Balcomb & Gerken, 2008) and it might be uniquely human (Metcalfe & Kober, 2005). Accordingly, there has been sharp interest in this cognitive capacity (e.g., Hertzog & Dunlosky, 2004; Koriat, Goldsmith, & Pansky, 2000; Koriat &

Levi-Sadot, 2001; Nelson & Dunlosky, 1991). Given the potential adaptive significance of metacognition, it would not be surprising to find that nonhuman animals share some form of metacognitive abilities, and many empirical studies and theoretical debates have emerged in recent years to address this question (e.g., Carruthers, 2008, 2009; Crystal & Foote, 2009; Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009; Smith, 2009; Kornell, 2009, 2013; Le Pelley, 2012; Smith, Beran, Couchman, & Coutinho, 2008).

The uncertainty-monitoring task is a prevalent paradigm for assessing metacognition in nonhuman animals. Subjects typically make a dichotomous classification of a stimulus presented on the screen (e.g., determining if a pixelated square is *sparse* or *dense*). If correct, subjects hear a melodic tone and receive a food pellet. If not, a timeout begins and the screen is blank until the next trial. Sometimes a third response option - typically a ? on the screen - allows subjects to opt out of a trial (all stimuli disappear from the screen) with no direct reward (e.g., the food reward) and no direct punishment (e.g., the penalty timeout for an error). The next trial simply begins. Rhesus macaques (*Macaca mulatta*) often selectively opt out of more objectively more difficult trials, as defined by proximity to the objective midpoint of the continuum. They choose the uncertainty response most often for exactly those trials where they are at greatest risk for making an error (see Smith, 2009). They do this in a wide variety of tasks and circumstances, including perceptual classification tasks with deferred feedback (Smith, Beran, Redford, & Washburn, 2006), same-different assessments (Shields, Smith, & Washburn, 1997), numerical judgments (Beran, Smith, Redford, & Washburn, 2006), and when they perform as many as four different tasks at the same time (Smith, Redford, Beran, & Washburn, 2010). However, there is not a consensus about the mechanisms underlying these performance patterns, and many empirical reports have been conducted to better understand the mechanism(s) underlying uncertainty responding by rhesus monkeys. Often, human participants are given the same tasks as monkeys, and perform similarly in terms of their use of the uncertainty response (and even say that they use it when they are not sure how to otherwise respond), but this does not mean that the same mechanism - metacognition - underlies the performance of both species. Thus, the debate continues about the proper interpretation of data from this test paradigm (Carruthers, 2008, 2009; Crystal & Foote, 2009; Hampton, 2009; Kornell, 2009, 2013; Le Pelley, 2012; Jozefowicz et al., 2009; Smith, 2009; Smith et al., 2008).

Capuchin monkeys (*Cebus apella*) - a New World primate species - do not show the use of the uncertainty response that is clearly expressed by apes and macaques (e.g., Beran, Smith, Coutinho, Couchman, & Boomer, 2009). The latter species commonly select the uncertainty response and use it most often when trials are objectively most difficult. Capuchin monkeys rarely do so, even when there are extremely long penalty timeouts for incorrect classification responses (e.g., Beran et al., 2009). This might suggest that capuchin monkeys lack whatever metacognitive monitoring capacity exists for rhesus monkeys. Recently, however, we have shown that capuchins can make appropriate uncertainty responses if one sharply reduces their probability of being correct by chance guessing (Beran, Perdue, & Smith, 2014). This suggests a species difference in risk tolerance between the capuchins and macaques. And if there are differences across species in the sensitivity to contingencies like risk, there might also be differences across species in the sensitivity to other contingencies that are present in the uncertainty monitoring task.

More specifically, an important aspect of the uncertainty-monitoring paradigm is that use of the uncertainty response need not be positively reinforced (though in some studies it has been, see discussion in Smith et al., 2008). Generally, in research with macaques, a food reward is not delivered for making that response. Instead, the uncertainty response avoids a potential timeout, and moves to the next trial (which may be harder or easier). But, choice of the uncertainty response also eliminates any chance for food reward. This is a critical aspect to its function, but one that is also sometimes forgotten. Monkeys must know what trials are best for using the uncertainty responses (e.g., those with high degrees of risk of error) rather than to use it indiscriminately because then many easy trials that should produce food reward will be lost.

With regard to avoiding timeouts, the function of the uncertainty response can be seen as providing the subject with negative reinforcement because its selection avoids a potentially aversive stimulus (the long timeout that is given only for incorrect responses). One possibility is that avoidance learning could be shaping the use of the uncertainty response as follows. Some stimuli (i.e., the most difficult stimuli near the breakpoint of the discrimination) could serve as discriminative stimuli indicating that the uncertainty response will lead to the avoidance of an otherwise likely timeout if a primary response is made. Thus, over time, the behavior of selecting the uncertainty response might increase in the presence of those stimuli because of negative reinforcement (choice of the uncertainty response takes away the negative timeout stimulus). This type of learning has been demonstrated in many species and, if this is in fact occurring with the perceptual judgment task, perhaps capuchin monkeys do not use uncertainty responses not because they cannot monitor their own capacity for classifying stimuli, but because they are not as sensitive to negative reinforcement relative to positive reinforcement, and so the value of the uncertainty response is not the same for them as it is for rhesus monkeys. Thus, guessing incorrectly (but, with a 50% chance of still being correct given that there are only two classes of stimulus denseness) is viewed as a better outcome than avoiding a timeout but with a 100% cost of also eliminating any chance for a food reward. As noted earlier, there is some evidence that capuchin monkeys may be more risk tolerant than macaques in these contexts (Beran et al., 2014), and so we attempted to assess this aspect of the temporal contingencies of the uncertainty monitoring task.

We presented capuchin monkeys with a new variation of the dense-sparse task to explore whether differential sensitivity to temporal contingencies might account for the previous failure of these monkeys to use an uncertainty response. We used an uncertainty-monitoring task in which correctly made primary discrimination responses (i.e., sparse/dense responses) *and* the uncertainty response all resulted in the same kind of contingency. Monkeys still performed a box density classification, and could still label a box as sparsely pixelated, densely pixelated, or could select the uncertainty response. Now, however, for the first time these three response classes generated the same qualitative outcome, but to quantitatively different degrees. In all cases, the outcomes of these choices dictated the temporal delay until a second task was presented, and it was this task that afforded the monkeys the opportunity to earn potential food rewards. This second task was a matching-to-sample (MTS) task, in which monkeys viewed a sample on the screen and then selected a match choice that was identical to the sample from an array of possible matches. If they selected

correctly *then and only then* did they receive a food reward. For the density discrimination task there was no chance to receive food rewards. Rather, correct primary classification (Dense/Sparse) led to nearly immediate presentation of the MTS task. Incorrect classifications led to a long temporal delay before the MTS trial was presented, and the uncertainty response led to an intermediate delay until presentation of the MTS trial.

It is critical to note that appearance of the sample for the MTS trial was not differentially probable on the basis of the density classification. It always appeared whether the monkey was correct, incorrect, or chose the uncertainty response, and thus could not operate as a secondary reinforcer that selectively reinforced certain choices in the density discrimination task. And, it is critical to note that the monkeys then still had to accurately perform the MTS trial to get any food reward. If they were wrong, they sat through another timeout. Thus, reward was entirely contingent on performing the MTS task well, and had nothing to do with performance on the density discrimination task. The only impetus for doing well at the density classification (and perhaps learning when to use the uncertainty response) was to minimize the delay that otherwise existed before the MTS trial could be performed.

If capuchin monkeys were unable to learn and perform even the basic density discrimination task under these newly established contingencies, it would suggest that differential sensitivity to negative reinforcement (i.e., removing the time delay until the MTS trial was presented) is a viable explanation for capuchin monkey responses patterns, and one that is different from a claim that these monkeys lack access to metacognitive signals that underlie uncertainty monitoring. In our view, this is an important issue in further understanding the nature of the uncertainty response as it is experienced by capuchin monkeys, and in better defining the role of stimulus control in these kinds of tasks. However, if capuchin monkeys learned the basic sparse-dense discrimination proficiently, but still did not use the uncertainty response, this would allow us to discount the hypothesis that the different behavioral profiles seen in the two species in uncertainty-monitoring tasks is the result of capuchin monkeys lacking sensitivity to the value of minimizing temporal delays in tasks that also produce food rewards.

Experiment 1

Sparse/dense trials involve the presentation of a pixelated box in the middle of the computer screen in which the number of pixels could range from sparsely to densely pixelated on a continuum. Subjects select a *S*, *D*, or *?* icon on the screen to classify the pixelated box (see method section for more details). We presented capuchin monkeys with a task in which a trial began with this sparse/dense judgment before monkeys were presented with a matching-to-sample (MTS) trial, and it was only the MTS trial that afforded the monkey a chance to obtain food reward. This modification from the standard task was made to ensure that the responses on the sparse/dense task were

never directly positively reinforced. Rather, they led to a separate, distinct task that could potentially yield food reinforcement depending on the accuracy of the response.

For the sparse/dense judgment task, three response options were available on the screen during each trial (*S* for sparse stimuli, *D* for dense stimuli, and ? as the uncertainty response). A correct response in classifying a stimulus as dense or sparse led to the immediate presentation of a MTS task but no food reward or other tangible positive reinforcer. An incorrect sparse or dense response led to a 30 s delay before presentation of the MTS task. The ? response led to a 6 s delay before the onset of the MTS task. Thus, correct responses on the sparse/dense judgment allowed the subject to entirely avoid longer delays to another task, but were never directly positively reinforced. Similarly, the ? response avoided a longer delay that could have occurred with an incorrect response but still added a longer delay than would have occurred for a correct response. The uncertainty response also was not directly reinforced.

Once the MTS trial was presented, a correct matching response led to a food reward whereas an incorrect response led to another timeout period before the next sparse/dense trial was presented. Thus, the only way to earn positive reinforcement in the form of food reward was through correct matching. It is critical to note that this MTS trial always appeared. Thus, the presence of a sample in the middle of the screen could not operate as any kind of conditioned reinforcer of responses made to specific density boxes in the first component of the task, because the presence of a sample always happened.

Given capuchin monkeys' limited use of the uncertainty response in past tasks with this design (e.g., Beran et al., 2009), their performance in this experiment allowed insight into what mechanisms might underlie uncertainty responding in this type of task. More specifically, this experiment presented an opportunity to test the alternative hypothesis that capuchin monkeys previously failed to produce adaptive patterns of uncertainty responding in this density discrimination task because they are less sensitive to the temporal contingencies of these tasks than other species. There were three possible outcomes of interest:

1. Capuchin monkeys fail to learn to use any of the discrimination responses. This would include failing to maintain sparse, dense, or uncertainty responses appropriately.

This finding would suggest that capuchin monkeys are not proficient at learning in general through the use of presenting or removing time delays that slow eventual access to food rewards. This would also suggest that some uncertainty-monitoring paradigms used with macaques may not be fair tests when used with capuchins because of this differential learning sensitivity.

2. Capuchin monkeys succeed in using all three responses appropriately (sparse, dense, uncertainty).

This finding would suggest that capuchin monkeys are sensitive to the negative contingencies (removal of time delays), and that the uncertainty response may have failed to emerge in capuchins in previous studies because the primary and uncertainty responses were learned under different contingencies (i.e., positive reinforcement for the dense and sparse responses, negative reinforcement for the uncertainty response).

This would suggest that uncertainty monitoring in capuchin monkeys requires an equivalent reinforcement (and punishment) structure across the response classes.

3. Capuchin monkeys successfully use the primary discrimination responses under negative contingencies, but do not use the uncertainty response when appropriate.

This finding would suggest that capuchin monkeys are sensitive to negative contingencies in this kind of perceptual judgment task because they can make responses that solely operate to remove time delays. When the temporal contingencies are all placed on the same plane within the task, the monkeys can judge density, but still do not anticipate and respond to difficult trials that risk making errors. This finding would argue against the hypothesis that species differences in uncertainty responding can be explained by species differences in sensitivity to negative reinforcement, but would leave open the question of whether capuchin monkeys lack metacognitive monitoring abilities or are constrained by other aspects of the task demands that are not directly relevant to metacognitive monitoring.

Method

Subjects. Ten adult capuchin monkeys between the ages of 5 and 13 years were tested. Five of these monkeys had participated in the previous Beran et al. (2009) study (three males - Logan, Liam, and Griffin, and two females - Wren and Nala) whereas the other five monkeys were naïve with regard to this computerized task (two males - Nkima and Drella, and three females - Gambit, Widget, and Lily). Although these monkeys were naïve to the sparse/dense task, they already had learned and were highly proficient with MTS tasks, how to respond to computer-generated stimuli, and how to engage those tasks using the joystick. All monkeys were housed in social groups of six individuals, and these monkeys spent the majority of the days in these social groups with access to outdoor yards and continuous access to water. Monkeys voluntarily separated into individual test boxes that allowed them to access a dedicated computer system for assessing various cognitive capacities (e.g., see Beran, 2008; Beran, Evans, Klein, & Einstein, 2012; Beran & Parrish, 2012; Smith et al., 2012). These test sessions typically lasted between 2 and 4 hours, and then monkeys returned to the larger social group.

Apparatus. The monkeys were tested using the Language Research Center's Computerized Test System. This system consisted of a personal computer, digital joystick, color monitor, and pellet dispenser (Evans, Beran, Chan, Klein, & Menzel, 2008; Richardson, Washburn, Hopkins, Savage-Rumbaugh, & Rumbaugh, 1990). Monkeys manipulated the joystick with their hands to produce isomorphic movements of a small cursor somewhere on the computer screen. Contacting stimuli with the cursor sometimes resulted in the delivery of 45-mg banana-flavored chow pellets (Bio-Serv, Frenchtown, NJ) via a pellet dispenser that was connected to the computer. The task program was written in Visual Basic 6.0.

Design and procedure. We presented monkeys with a task in which a trial began with a sparse/dense judgment that monkeys could make or could avoid by choosing the uncertainty response. A correct sparse or dense response led to the immediate presentation of an MTS task. An incorrect sparse or dense response led to a 30 s delay before presentation of the MTS task. The uncertainty response led to a 6 s delay before presentation of the MTS task. Thus, correct responses on the sparse/dense judgment allowed the subject to avoid longer delays to another task, but were never directly positively reinforced. Similarly, the uncertainty response avoided the long delay of an incorrect response but still gave a longer delay than a correct response. It was not directly reinforced.

Density discrimination. On each trial, monkeys saw a 185 X 185-pixel box in the screen's top center. The box was filled with a varying number of randomly placed white pixels presented on a black background. Sixty stimulus levels could be presented (Levels 1-60). Each level's pixel count was given by the formula $\text{pixels} = \text{round}(\text{base pixels} \times 1.036^{\text{Level}})$ where the base pixel number was 400. This formula gave the continuum a logarithmic character, with each density step a constant percentage increase in pixels and not a constant absolute increase in pixels. Half of the trials were objectively defined by the program as being sparse (Levels 1-30) and half were dense (Levels 31-60). Approximately 30% of the trials were restricted to only the central third of the range (Levels 21-40) whereas the other 70% were sampled

from the full continuum (Levels 1-60; note that Level 21-40 trials also were possible as part of this second sampling method).

Capuchin monkeys touched an *S* or *D* icon located at left center and right center of the screen to classify the pixelated box as sparse or dense, respectively. Correct responses immediately led to the presentation of the MTS trial (see description below). When an *S* or *D* response was made incorrectly, there was a 30 s interval with a blank screen, and then the MTS trial was presented. Choice of the uncertainty response (the ? at the bottom center) led to a 6 s blank screen followed by presentation of the MTS trial. Thus, there was no positive reinforcement presented for correct responses. Rather, those responses were negatively reinforced by removing any delay to the MTS part of the trial. Incorrect responses were positively punished (as in past research using this paradigm), and uncertainty responses operated to alleviate some of the delay to the MTS part of the trial (i.e., were negatively reinforced but to a smaller degree than correct primary responses).

Matching-to-sample (MTS) task. In the MTS task, a sample image first appeared in the center of the screen. This image was randomly selected from a large library of distinctive clip art images. Monkeys made an observing response to this sample by moving the cursor into contact with it. Then, one identical matching stimulus and three non-matching comparison stimuli appeared randomly assigned in the four corners of the computer screen. Monkeys had to choose the comparison stimulus that was identical to the sample. There was no memory demand as the sample remained onscreen with the match choices, and thus the task was very easy for the monkeys (they typically performed at 90-100% accuracy on these trials). Correct responses led to a melodic tone and delivery of a food pellet whereas incorrect responses led to a buzz tone and a 30 s timeout during which the screen remained blank. Following either of these outcomes, a 1 s inter-trial interval occurred, and then the next density discrimination was presented.

Because monkeys worked at their own pace, they completed variable numbers of trials within experimental sessions and across the whole experiment. We first examined performance when the trial count for a monkey was approximately 4,000 trials. If subjects only used one response during the density discrimination task (i.e., chose the sparse stimulus 80% or more of the trials, or conversely selected the dense stimulus on 80% or more of trials) after these approximately 4,000 trials, participation in the task was stopped. All other subjects continued in the experiment until all monkeys had completed approximately 10,000 trials or more. Some monkeys exceeded this number substantially because they tended to perform many more trials per session than other monkeys. One monkey, Wren, was not regularly engaging this task during the time period of this study, and so we discontinued testing her given this low motivation level to engage the task but still included her data in analyses. Thus, her lower data count reflects her choice to not participate as often as the other monkeys. Trial counts are provided in Table 1.

Table 1

Trial counts for Experiments 1 and 2

Monkey Name	Trials	
	Experiment 1	Experiment 2
Capuchin Monkeys		
Drella	4,138	19,505
Gambit	6,510	9,737
Lily	8,528	5,278
Griffin	22,663	6,258
Wren	3,991	2,553
Logan	26,497	9,140
Nala	9,911	9,975
Nkima	9,975	12,410
Liam	9,974	9,977
Widget	9,661	---
Gabe	---	9,560

Data analysis. Group level (see Figure 1) and individual (see Figure 2) plots of the sparse, dense and uncertainty responses were made. For the purpose of statistical analysis, the stimulus levels were collapsed into bins. To approximate an equal number of trials in each bin, 2-level bins were created for the middle region (because this was oversampled in the task) and 4-level bins were created for the ends. For the sparse and dense response, the percent correct was calculated for each bin. For the uncertainty

response, the overall percent of trials in which it was used was calculated for each bin. Subjects that showed an 80% or greater biased selection of one response (Drella, Gambit and Lily, see Figure 2) were excluded from the analysis. We used three separate repeated-measures analysis of variance (ANOVA) tests to assess the effect of bin level on the sparse, dense and uncertainty response. For a more conservative statistical approach, we used the Huyhn-Feldt correction (Huyhn & Feldt, 1976).

Results

Subjects were highly proficient in the MTS task - typically ranging between 90% and 100% correct. Three (out of 10) monkeys (Drella, Gambit, Lily) did not succeed on the density discrimination task and primarily (80% or more) used one of the response options for the majority of trial levels. Participation by those monkeys was discontinued after approximately 4,000 trials. These three monkeys were three of the five that were naïve to the sparse/dense discrimination task at the onset of the experiment. In general, not all monkeys used the uncertainty response in previous experimental tasks (e.g., Smith et al., 2006; Smith et al., 2010), suggesting that this is a challenging task and it is not entirely surprising that some naïve monkeys did not initially learn the task. The other seven monkeys, including two naïve subjects, showed primary discrimination curves that nicely approximated the expected responses based on the perceptual continuums. Specifically, as supported by visual inspection of the data shown in Figure 1, we expected sparse responses to decrease, and dense responses to increase, as the level increased. There was a significant effect of bin on proportion of sparse responses, $F(3.7, 22.3) = 366.73, p < 0.001$, and dense responses, $F(3.7, 22.3) = 366.73, p < 0.001$. However, bin level did not have any effect on the use of the uncertainty response, $F(4.2, 25.3) = 2.05, p = 0.115$ (Figure 1), suggesting that subjects did not use the uncertainty response in a manner that suggested uncertainty monitoring, with an increase in UR use as objective trial difficulty increased. These monkeys all showed clear discrimination of the sparse and dense regions even though the responses in classifying those stimuli as sparse or dense were never directly reinforced with food pellets in this experiment.

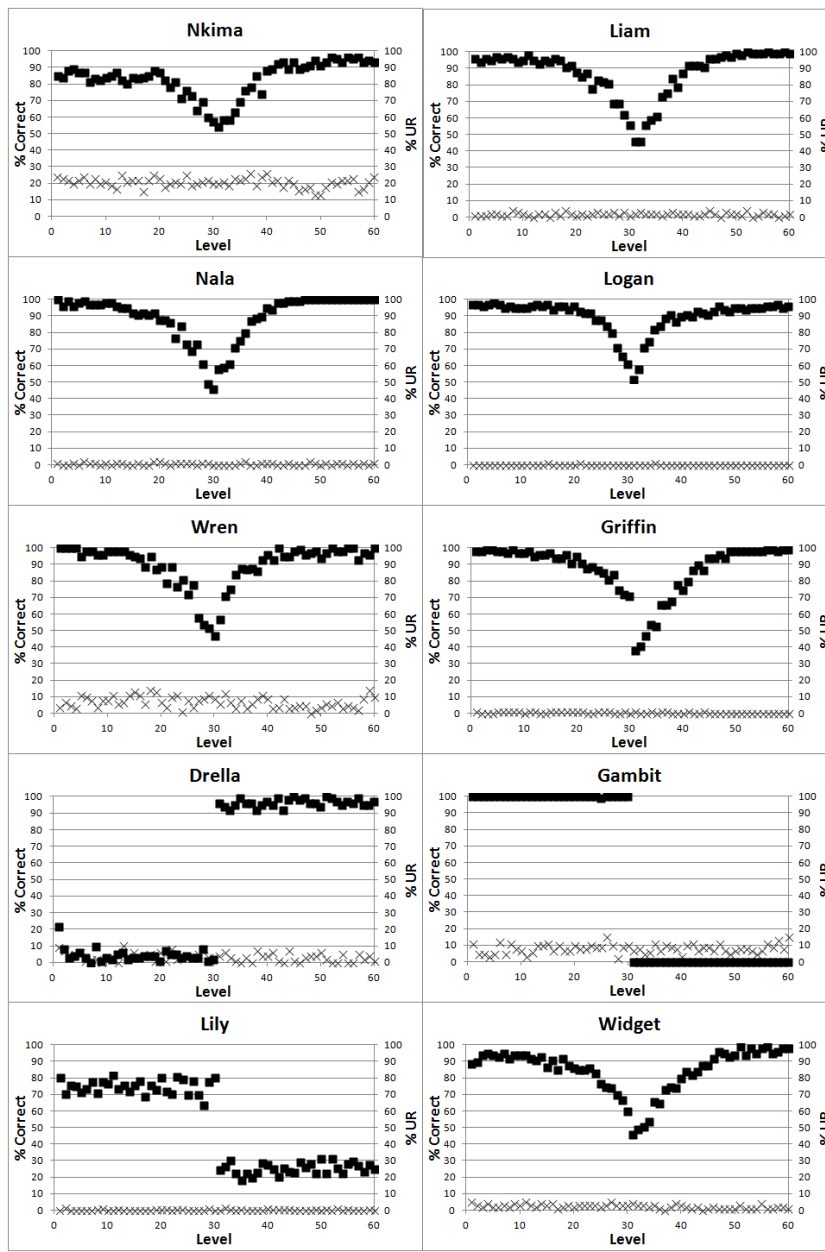
Figure 1. Average proportion of sparse, dense and uncertainty responses across bin levels.

In terms of our potential outcomes, the results are mostly in line with the third potential outcome that stated that monkeys could learn the primary discrimination in a perceptual judgment task via negative reinforcement, but not integrate proficient use of the uncertainty response. This finding suggests that in this type of perceptual judgment task, capuchin monkeys are capable of learning the appropriate task responses via negative reinforcement contingencies for those responses, but they still do not produce adaptive uncertainty monitoring curves even when all response classes are similar in nature. We can reject the idea that capuchins are not capable of learning through negative reinforcement (potential Outcome 1) or that they learn all response classes when placed under the same contingencies (potential Outcome 2).

Figure 2. Capuchin monkey performance (% Correct - black squares; % Uncertainty Response - gray hatched marks) in Experiment 1 level.

Discussion

This was designed so primary discrimination yielded the of outcome (determining the another task) as uncertainty. The majority of monkeys were learning and a perceptual discrimination these. Unlike previous they were never rewarded for a sparse or dense. Instead, a response timeout to task that did positive reinforcement responses.



Experiment 1 level.

experiment that

responses same type

delay to the response. capuchin capable of performing

task under conditions. studies, positively correct response. correct avoided a another provide

for correct Given these

contingencies, seven out of 10 monkeys discriminated sparsely and densely pixelated images from one another.

However, these monkeys still did not use the uncertainty response adaptively to avoid the most difficult trials. To the capuchins, this third response in the task was somehow meaningfully different, even though all three responses acted only to determine the temporal delay until the MTS task was presented. Thus, one cannot attribute the failure of capuchins to respond uncertain in the previous study (Beran et al., 2009) to their insensitivity to avoiding time outs. For most monkeys, their sensitivity was fine when it came to perceptual-discrimination responses. Instead, capuchins exhibited a selective failure to use the uncertainty response in particular. One must seek another behavioral or psychological explanation for this selective failure, and we return to this issue in the General Discussion.

Crucially, we tested five naïve animals that had not been in the Beran et al. (2009) study, although one of these monkeys (Drella) had some experience classifying pixel boxes (but not using uncertainty responses) as part of a pilot study for the Beran et al. (2009) study. The other four monkeys had no experience classifying pixel boxes as sparse or dense. Although Drella, Lily, and Gambit showed strong response biases and seemed to have some difficulty in learning the classification categories under these contingencies, Widget and Nkima did learn the primary discrimination, but these monkeys failed to show any proficient uncertainty responding.

The perceptual discriminations used in Experiment 1 were objectively easier than those used in previous experiments (e.g., Beran et al., 2009). The changes in density level to level were larger than typically used. This was intentionally done to motivate participation in a task that did not in and of itself yield positive reinforcement. However, it is possible that the capuchin monkeys opted not to use the uncertainty response because the sparse and dense responses were subjectively easy to make (even if they were not always accurate), and this could account for the pattern of uncertainty responding in Experiment 1 (E1). Experiment 2 (E2) addressed this issue.

Experiment 2

Method

Subjects and apparatus. Nine of the same capuchin monkeys were tested as in Experiment 1. We added one new male (Gabe) and removed Widget from Experiment 2 due to her limited availability during the project. Even monkeys that were excluded from E1 analyses based on performance patterns were included again at the start of this new phase. The apparatus was identical to that in Experiment 1.

Design and procedure. All aspects of the task were the same except for the following:

Although 60 stimulus levels again could be presented (Levels 1-60), each level's randomly placed pixel count now was given by the formula $\text{pixels} = \text{round}(\text{base pixels} \times 1.018^{\text{level}})$ where the base pixel number was 689. This formula again gave the continuum a logarithmic character, with each density step a constant percentage increase in pixels and not a constant absolute increase in pixels. This change in base pixel number and multiplier value made the task objectively more difficult as each level was closer in comparison to its neighbors than had been the case in Experiment 1. Monkeys again performed variable numbers of trials as a result of individual differences in productivity within individual test sessions. Trial counts for each monkey are provided in Table 1.

Results

All subjects were highly proficient on the MTS task - typically ranging between 90% and 100% correct.

Three out of 10 capuchin monkeys (Gambit, Lily and Nkima) did not succeed on the primary discrimination and were excluded from the statistical analysis. The other seven monkeys showed primary discrimination curves that nicely approximated the expected responses (as detailed in E1) to sparse stimuli, $F(1.4, 8.2) = 58.59, p < 0.001$, and dense stimuli, $F(1.4, 8.2) = 58.59, p < 0.001$; however, they did not use the uncertainty response, $F(3.9, 23.6) = 1.06, p = 0.398$, in a selective manner (see Figures 1 and 3).

We also assessed whether the objective difficulty changes made to the program led to differences in performance. For the animals that learned the primary discrimination in both experiments, performance was significantly better in E1 ($M = 88.79, SD = 1.04$) than in E2 ($M = 80.89, SD = 3.72$), $t(4) = 6.08, p = 0.004$.

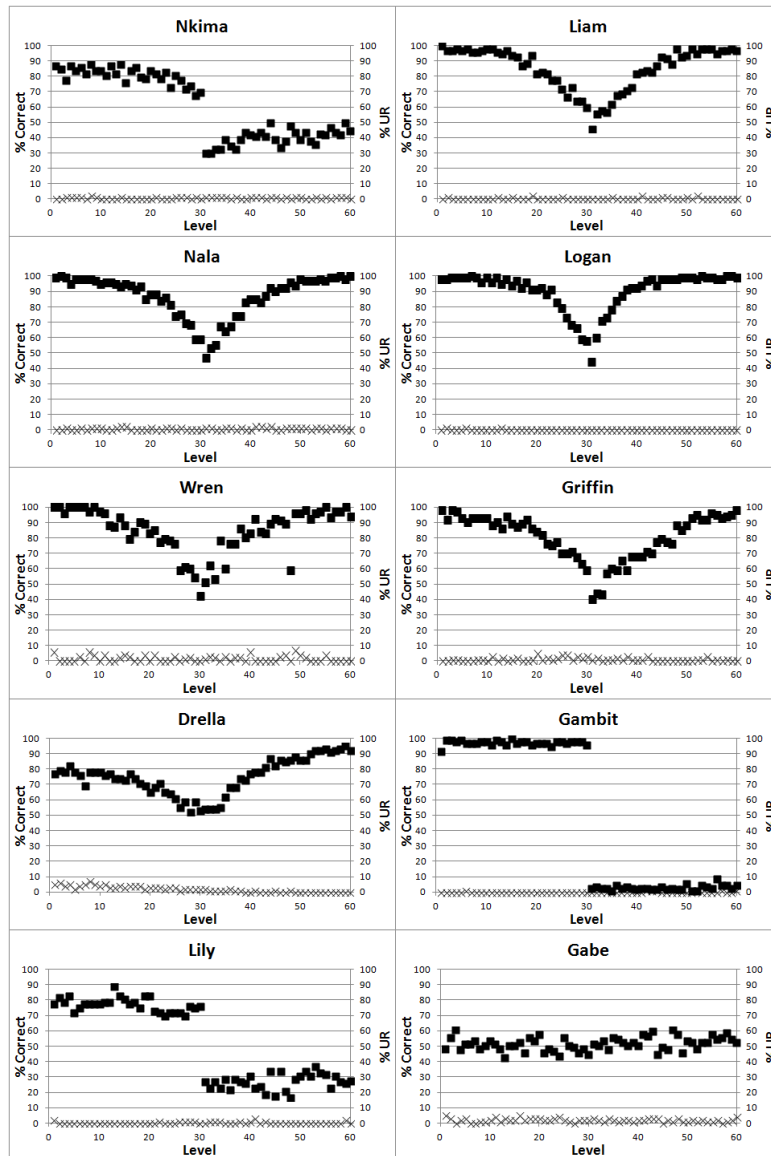


Figure 3. Capuchin monkey performance (% Correct - black squares; % UR - gray hatched marks) in Experiment 2 as a function of trial level.

Discussion

Experiment 2 systematically tightened the psychophysical discrimination, adding in more objective task difficulty by presenting more trials that were closer to the Sparse-Dense divide than in Experiment 1, plausibly increasing capuchins' motivation to use the uncertainty response adaptively to decline the most difficult trials (or, interpreted associatively: to avoid the most error-causing stimuli). The density discrimination part of the experiment was still run with no positive reinforcement for the sparse, uncertain, and dense responses. The result was the same as in Experiment 1. Capuchin monkeys were capable of performing the more difficult perceptual discrimination task and still did not use the UR. However, these monkeys still did not respond by using the uncertainty response to avoid longer delays on the most difficult trials. One cannot explain this result using the ease of the task in Experiment 1.

General Discussion

The findings of the present work suggest that a potential lack of, or limitation in learning via negative reinforcement is *not* a likely reason that capuchin monkeys do not succeed on uncertainty responding tasks. Thus, the data support our third outcome of interest described earlier while not supporting the first two outcomes that were noted. There has been the fairly consistent finding that capuchin monkeys fail to produce the adaptive response patterns of rhesus monkeys in uncertainty monitoring tasks as well as other tasks that are proposed to assess metacognition across species. Rhesus monkeys efficiently seek information about where food might be hidden (Basile, Hampton, Suomi, & Murray, 2009), but capuchin monkeys do not (Paukner, Anderson, & Fujita, 2006). Rhesus monkeys search for specific kinds of needed information during computerized memory tests whereas capuchin monkeys do not (Beran & Smith, 2011). And in tasks such as the one used in this experiment, macaques use the uncertainty response most often when they encounter difficult discriminations, whereas capuchin monkeys do not (e.g., Beran et al., 2009; but see Beran et al., 2014). One possible explanation for the apparent species difference on these tasks is that capuchins indeed lack metacognitive abilities present in some other primate species. However, it is critical to assess the possibility that researchers have yet to determine the correct situations in which such capacities might manifest in capuchin monkeys. Failures on the existing tasks should be carefully assessed to determine if an explanation other than lack of metacognition might explain differences in performance.

Typically, in uncertainty monitoring tasks monkeys are rewarded for correct classifications (positive reinforcement) and punished for incorrect classifications, but use of the uncertainty response simply moves to the next trial. This simplicity, however, does not necessarily extend to a straightforward interpretation of exactly what an uncertainty response is for the subject. Here, we have addressed the concern that it might operate as a form of negative reinforcement in the sense that if the animal was about to suffer an error, and receive a penalty, that penalty was avoided by making the uncertainty response. Below, we discuss this possibility in more detail. However, it is critical to remind the reader that uncertainty responses also *cost* the subject the chance at a reward. There is no way to receive positive outcomes (food pellets) when using an uncertainty response that simply erases the trial. Thus, the uncertainty response perhaps causes the combined feelings of avoiding a possible timeout and giving up a possible reward – and what the sum of the two motivations is for the subject is unknown, particularly for nonverbal subjects.

Furthermore, the concern that capuchin performance might differ as a function of positional biases or differential sensitivity to delay has been addressed in previous work. Capuchin monkeys did not use the UR in Beran et al. (2009), but they did frequently and appropriately use the Middle response that was in the same location on the screen as the UR was located, but in a different trial type. And, some monkeys did use the UR located at the bottom of the screen in Beran et al. (2014). Thus, we can discount a bias to not move downward in discrimination tasks, and we note that monkeys will select the ? stimulus in some contexts. Capuchin monkeys are as sensitive to different delay levels as other primates. For example, we have conducted numerous delay discounting tasks with these specific monkeys, and they showed consistent within-subjects degrees

of delay discounting that also were similar to those seen in rhesus monkeys (e.g., Evans, Perdue, Parrish, & Beran, 2014).

The primary goal of this article was to explore whether subjects could learn a psychophysical discrimination task when the outcome contingencies for all three response classes were qualitatively similar (they all impacted delay to the MTS task, which was the only part of the experiment in which food reward could be earned). This discrimination task was not designed to promote the strongest and clearest use of the uncertainty response, but rather to contrast performance on this task with that of previous assessments with capuchin monkeys. In the current task, the uncertainty response actually led to a timeout period before the MTS trial appeared. The uncertainty response here served more as a “plea bargain” response that would allow subjects to endure a shorter timeout than an incorrect response would yield. The majority of the capuchins learned and maintained the primary classifications (sparse and dense) in this task, but still never appropriately used the uncertainty response. Capuchin monkeys were sensitive to temporal contingencies that determined how soon the next trial can appear and be completed. However, the monkeys only were successful in making the sparse and dense responses under these contingencies, whereas the uncertainty response, even when equated to the other responses in terms of its qualitative effect to dictate the delay interval, remained unused or used without any clear relation to objective trial difficulty. Thus, the search for uncertainty responding in capuchin monkeys remains ongoing, but one potential non-metacognitive reason for differential performance of this species compared to macaques now can be dismissed.

References

- Basile, B. M., Hampton, R. R., Suomi, S. J., & Murray, E. A. (2009). An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition*, *12*, 169-180.
- Beran, M. J. (2008). Capuchin monkeys (*Cebus apella*) succeed in a test of quantity conservation. *Animal Cognition*, *11*, 109-116.
- Beran, M. J., Evans, T. A., Klein, E. D., & Einstein, G. O. (2012). Rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*) remember future responses in a computerized task. *Journal of Experimental Psychology: Animal Behavior Processes*, *38*, 233-243.
- Beran, M. J., & Parrish, A. E. (2012). Sequential responding and planning in capuchin monkeys (*Cebus apella*). *Animal Cognition*, *15*, 1085-1094.
- Beran, M. J., Perdue, B. M., & Smith, J. D. (2014). What are my chances? Closing the gap in uncertainty monitoring between rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Learning and Cognition*, *404*, 303-316.
- Beran, M. J., & Smith, J. D. (2011). Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Cognition*, *120*, 90-105.
- Beran, M. J., Smith, J. D., Coutinho, M. V. C., Couchman, J. J., & Boomer, J. (2009). The psychological organization of “uncertainty” responses and “middle” responses: A dissociation in capuchin monkeys (*Cebus apella*). *Journal of Experimental Psychology: Animal Behavior Processes*, *35*, 371-381.

- Beran, M. J., Smith, J. D., Redford, J. S., & Washburn, D. A. (2006). Rhesus macaques (*Macaca mulatta*) monitor uncertainty during numerosity judgments. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*, 111-119.
- Call, J. (2010). Do apes know that they could be wrong? *Animal Cognition*, *13*, 689-700.
- Call, J., & Carpenter, M. (2001). Do apes and children know what they have seen? *Animal Cognition*, *4*, 207- 220.
- Carruthers, P. (2008). Meta-cognition in animals: a skeptical look. *Mind and Language*, *23*, 58-89.
- Carruthers, P. (2009). How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*, *32*, 121-182.
- Crystal, J. D., & Foote, A. L. (2009). Metacognition in animals: Trends and challenges. *Comparative Cognition and Behavior Reviews*, *4*, 54-55.
- Evans, T. A., Beran, M. J., Chan, B., Klein, E. D., & Menzel, C. R. (2008). An efficient computerized testing method for the capuchin monkey (*Cebus apella*): Adaptation of the LRC-CTS to a socially housed nonhuman primate species. *Behavior Research Methods*, *40*, 590-596.
- Evans, T. A., Perdue, B. M., Parrish, A. E., & Beran, M. J. (2014). Working and waiting for better rewards: Self-control in two monkey species (*Cebus apella* and *Macaca mulatta*). *Behavioural Processes*, *103*, 236-242.
- Fujita, K. (2009). Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition*, *12*, 575-585.
- Hampton, R. R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, *98*, 5359-5362.
- Hampton, R. R. (2009). Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews*, *4*, 17-28.
- Hertzog, C. & Dunlosky, J. (2004). Aging, metacognition, and cognitive control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (p. 215 - 251). San Diego, CA:Academic Press.
- Huynh, H., & Feldt, L. S. (1976). Estimation of the Box correction for degrees of freedom from sample data in randomised block and split-plot designs. *Journal of Educational Statistics*, *1*, 69-82.
- Jozefowicz, J., Staddon, J. E. R., & Cerutti, D. T. (2009). Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews*, *4*, 29-39.
- Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology*, *51*, 481-537.
- Koriat, A., & Levy Sadot, R. (2001). The combined contributions of the cue-familiarity and accessibility heuristics to feelings of knowing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*, 34-53.
- Kornell, N. (2009). Metacognition in humans and animals. *Current Directions in Psychological Science*, *18*, 11-15.
- Kornell, N. (2013). Where is the "meta" in animal metacognition? *Journal of Comparative Psychology*, *128*, 143-149. doi:10.1037/a0033444
- Le Pelley, M. E. (2012). Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *38*, 686-708.

- Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science, 2*, 267-270.
- Paukner, A., Anderson, J. R., & Fujita, K. (2006). Redundant food searches by capuchin monkeys (*Cebus apella*): A failure of metacognition? *Animal Cognition, 9*, 110-117.
- Richardson, W. K., Washburn, D. A., Hopkins, W. D., Savage-Rumbaugh, E. S., & Rumbaugh, D. M. (1990). The NASA/LRC Computerized Test System. *Behavior Research Methods, Instruments, and Computers, 22*, 127-131.
- Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain response by humans and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task. *Journal of Experimental Psychology: General, 126*, 147-164.
- Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences, 13*, 389-396.
- Smith, J. D., Beran, M. J., Couchman, J. J., & Coutinho, M. V. C. (2008). The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin and Review, 15*, 679-691.
- Smith, J. D., Beran, M. J., Redford, J. S., & Washburn, D. A. (2006). Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General, 135*, 282-297.
- Smith, J. D., Coutinho, M. V. C., Church, B. A., & Beran, M. J. (2013). Executive-attentional uncertainty responses by rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: General, 142*, 458-475.
- Smith, J. D., Crossley, M. J., Boomer, J., Church, B. A., Beran, M. J., & Ashby, F. G. (2012). Implicit and explicit category learning by capuchin monkeys (*Cebus apella*). *Journal of Comparative Psychology, 126*, 294-304.
- Smith, J. D., Redford, J. S., Beran, M. J., & Washburn, D. A. (2010). Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal Cognition, 13*, 93-101.

Financial conflict of interest: This research was funded by NIH grants HD061455 and HD060563 and by NSF grant BCS0956993.

Conflict of interest: No stated conflicts.

Submitted: November 5th, 2014
Revision submitted: March 8, 2015
Accepted: May 2nd, 2015