Published in final edited form as:

Cognition. 2015 September ; 142: 236–246. doi:10.1016/j.cognition.2015.05.023.

Go When You Know: Chimpanzees' Confidence Movements Reflect Their Responses in a Computerized Memory Task

Michael J. Beran,
Georgia State University

Bonnie M. Perdue,
Agnes Scott College

Sara E. Futch,
Wofford College

J. David Smith,
University at Buffalo, The State University of New York

Theodore A. Evans, and
Georgia State University

Audrey E. Parrish
Georgia State University

Abstract

Three chimpanzees performed a computerized memory task in which auditory feedback about the accuracy of each response was delayed. The delivery of food rewards for correct responses was also delayed and occurred in a separate location from the response. Crucially, if the chimpanzees did not move to the reward-delivery site before food was dispensed, the reward was lost and could not be recovered. Chimpanzees were significantly more likely to move to the dispenser on trials they had completed correctly than on those they had completed incorrectly, and these movements occurred before any external feedback about the outcome of their responses. Thus, chimpanzees moved (or not) on the basis of their confidence in their responses, and these *confidence movements* aligned closely with objective task performance. These untrained, spontaneous confidence judgments demonstrated that chimpanzees monitored their own states of knowing and not knowing and adjusted their behavior accordingly.

Correspondence concerning this article can be addressed to: Michael Beran, Language Research Center, Georgia State University, University Plaza, Atlanta, GA 30302. mjberan@yahoo.com.
Michael J. Beran, Language Research Center, Georgia State University; Bonnie M. Perdue, Department of Psychology, Agnes Scott College; Sara E. Futch, Department of Psychology, Wofford College; J. David Smith, Department of Psychology, University at Buffalo, The State University of New York; Theodore A. Evans, Language Research Center, Georgia State University; Audrey E. Parrish, Language Research Center and Department of Psychology, Georgia State University.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Keywords

confidence judgments; chimpanzees; memory; matching-to-sample; metacognition

Humans provide confidence judgments about their own actions, knowledge, and abilities. They do this in many circumstances (e.g., studying, test taking, game-show participation), and those judgments are often assumed to reflect humans' metacognition (Benjamin, Bjork, & Schwartz, 1998; Dunlosky & Bjork, 2008; Flavell, 1979; Koriat & Goldsmith, 1994; Kornell, 2009; Nelson, 1992; Schwartz, 2008). Metacognition is considered to be a sophisticated cognitive capacity and perhaps even uniquely human (Metcalf & Kober, 2005). It emerges later in human development than many other cognitive capacities (Balcomb & Gerken, 2008). It is often expressed through verbal reports, though it can also be expressed behaviorally (e.g., a shrug of the shoulders). It is of intense interest to many researchers working to understand how humans comprehend, learn, and remember (e.g., Bjork, Dunlosky, & Kornell, 2012; Koriat & Goldsmith, 1996; Kornell & Son, 2009; Nelson & Dunlosky, 1991; Reder, 1987). The question of whether nonhuman animals might show degrees of metacognition is also intensely debated based on empirical reports of uncertainty-monitoring tasks, information-seeking tasks, and other tasks given to animals which often show strongly isomorphic performances in animals and humans (Carruthers, 2008, 2009; Crystal, 2014; Crystal & Foote, 2009; Hampton, 2009; Jozefowicz, Staddon, & Cerutti, 2009; Kornell, 2009, 2014; Le Pelley, 2012; Smith, 2009; Smith, Beran, Couchman, & Coutinho, 2008). It is an important question as to whether animals share any of the features of humans' metacognition, because evidence for or against such capacities in animals could inform us about the nature of their consciousness and self-awareness, and could affect many theoretical debates within comparative psychology, philosophy, and cognitive science. However, this does not mean that claims for evidence of metacognition in animals require acceptance of the idea of self-aware consciousness across species that show such metacognition. It simply means that studies of animal metacognition can help advance the pursuit of a better understanding of whether animals *might* show some similarities with humans in terms of conscious experiences.

Metacognition, as we use the term, refers to executively-controlled decision making in which behavior arises independently of external stimulus control. Metacognition is different from forms of learning (ranging from associative learning to observational learning) and forms of cognition (such as classifying, remembering, or judging) in that metacognitive processes are those that use the inputs from learning and/or cognitive processes to generate additional response classes that are not directly related to those other processes. Metacognitive processes produce signals to act, to hesitate, to search for more information, to move forward cautiously or confidently, and so forth. The present paper focuses on one of the outputs of these monitoring processes – confidence. For example, imagine being shown a large pile of pennies on a table and being asked to rapidly answer the question “How many are there?” or “Are there more than three pennies there?” In both cases, the perceptual experience is the same, and the capacity for generating an estimate of the number of pennies exists. But, with limited time, the first question would be impossible to answer whereas the second is entirely possible, and most adults would hesitate to give a response to the first

question, or indicate their response was likely going to be wrong, whereas the second question was easy to answer, and confidence would be high. In this example, the monitoring of cognition (here, as a process of rapid enumeration) is very different depending on the question, and the metacognitive output would be quite different. If the person was first told that they had 10 minutes to answer the first question, however, the metacognitive process would generate a much different feeling of confidence, assuming the person knew they would have enough time to count all of the pennies and give an exact (and accurate) response.

In formal experimental settings, confidence judgments by humans can take different forms—numerical ratings, verbal declarations, and so forth. Animals cannot easily report their confidence in these ways. However, rhesus monkeys and rats can provide other responses that reflect the monitoring of uncertainty of their perceptual experiences (Beran, Smith, Redford, & Washburn, 2006; Foote & Crystal, 2007, 2012; Shields, Smith, & Washburn, 1997; Smith, Beran, Redford, & Washburn, 2006; Smith, Redford, Beran, & Washburn, 2010; Smith, Schull, Strote, McGee, Egnor, & Erb, 1995; Smith, Shields, Schull, & Washburn, 1997) and pigeons, rhesus monkeys, and orangutans provide responses that reflect monitoring of memory (Adams & Santi, 2011; Basile, Hampton, Suomi, & Murray, 2009; Fujita, 2009; Hampton, 2001; Hampton & Hampstead, 2006; Inman & Shettleworth, 1999; Smith, Shields, Allendoerfer, & Washburn, 1998; Suda-King, 2008; Suda-King, Bania, Stromberg, & Subiaul, 2013; Sutton & Shettleworth, 2008; Templer & Hampton, 2012). Rhesus monkeys, chimpanzees, orangutans, and pigeons have demonstrated the ability to search for needed information (Beran & Smith, 2011; Beran, Smith, & Perdue, 2013; Call, 2010; Call & Carpenter, 2001; Castro & Wasserman, 2013; Hampton, Zivin, & Murray, 2004; Iwasaki, Watanabe, & Fujita, 2013; Kirk, McMillan, & Roberts, 2014; Marsh & MacDonald, 2012a, 2012b; Roberts et al., 2009), and have even shown the capacity to make confidence judgments about the outcomes of already completed responses, although only with extensive training (Kornell, Son, & Terrace, 2007; Morgan, Kornell, Kornblum, & Terrace, 2014; Nakamura, Watanabe, Betsuyaku, & Fujita, 2011; Shields, Smith, Guttmanova, & Washburn, 2005).

For example, Kornell et al. (2007) reported that rhesus monkeys provided confidence judgments about responses they had just made by selecting one of two icons to bet high or to bet low on that response. The monkeys tended to bet high more often on trials that were correct and low more often on trials that were incorrect. Once trained on how these icons operated, the monkeys transferred this appropriate use of the icons to new tasks assessing different forms of perception and memory (see also Morgan et al. 2014; Nakamura et al., 2011). Shields et al. (2005) compared the performance of rhesus monkeys to that of humans in another task that required a betting response about the perceptual classification that had just been made. Although humans showed a more complicated confidence-rating pattern, the monkeys showed some similarities. In particular, they sometimes chose to take smaller or larger risks on the classification they had just made. These bets matched the general pattern of classification performance, with higher bets more often made for correctly completed classifications.

These confidence-judgment studies all depended on extensive task experience and response training. Possibly, therefore, the entrained confidence responses might reflect some associatively learned chained patterns of behavior rather than metacognitive states and judgments. Therefore, one goal of our experimental design was to find a way to measure confidence behaviorally that would depend little on task experience and response training. Another aspect of our experimental design was to provide a behavioral expression of confidence that would reflect metacognitive monitoring while being intuitive to the subject and appealing/attractive in its use and function. Shields et al. (2005) noted that:

A burden falls on the experimenter to arrange carefully the methods of a confidence-monitoring experiment so that animal participants will be strongly motivated to adopt a metacognitive strategy if they are able. This can be a complex matter to judge in planning an experiment, and it may be a common problem for task grammars to motivate a metacognitive strategy too weakly for the animal to struggle to adopt it (Shields et al., 2005, p. 184).

One way to approach these experimental goals is to think about the natural behaviors that might reflect confidence in animals. For example, a leap from one branch to another by a monkey without hesitation could suggest that the monkey was confident it could clear the gap between those branches. Likewise, a movement through space to a food source could be an indication that the animal was confident that food would be at that location when it arrived. These examples illustrate our experimental approach. We designed an experiment in which the confidence judgments expressed by chimpanzees were natural—they were movements toward the location of food items. Particularly, the chimpanzees moved towards the location in which food would be presented if it had been earned by a correct response in the current task. This response required little or no training, meeting one of our empirical goals. It was intuitive and transparent and valuable to the animal, meeting another goal. It also provided us with a behavioral expression of confidence that was discrete, and thus easy to score.

We gave chimpanzees the opportunity to spontaneously indicate their confidence in this way within the context of obtaining rewards from a computer task. We presented the chimpanzees with a computerized matching-to-sample (MTS) task in which they would make correct or incorrect responses, depending on the reliability of their memory. These chimpanzees were experienced in performing such computerized tests and in receiving food rewards for correct responses. But such rewards had always been given either by an experimenter seated near the testing area, or by an automated dispenser that delivered food reward directly to the chimpanzee where it sat to work on the task. In the present experiments, food rewards were delivered distantly, requiring the chimpanzees to leave the computer apparatus and walk to where food would be dispensed if their response had been made correctly (Figure 1).

The chimpanzees also were accustomed to receiving their reward immediately after making a correct response. Now, however, the food delivery was delayed, so that chimpanzees would have time to move to the dispensing location before food was delivered. If they were not at that location in time, any delivered food was lost and was not recoverable because it fell away from the enclosure and remained on the floor outside the enclosure until it was

discarded after the test session. This created a situation in which chimpanzees could express clear, untrained confidence judgments about their own task performance through their movements in space. Specifically, chimpanzees—if they have the ability to assess their own memory and their correctness on each trial—should leave the test area and walk over to the dispensing area prior to reward delivery more often on trials that were correctly completed than on those not correctly completed. Given previous evidence of chimpanzee metacognition in a different (information-seeking) paradigm (Beran et al., 2013), we hypothesized that chimpanzees would exhibit these confidence judgments in the current study.

Experiment 1

Participants

Two chimpanzees were tested: Lana (43 year old female) and Sherman (41 year old male). Both chimpanzees had long testing histories using this computerized apparatus (e.g., Beran, 2006; Beran, Pate, Washburn, & Rumbaugh, 2004; Beran & Rumbaugh, 2001; Beran & Washburn, 2002), including past experience matching visual sample stimuli to correct match options in MTS tasks. Lana and Sherman also had learned to associate symbols called lexigrams with real world referents as part of language-acquisition projects (Rumbaugh, 1977; Savage-Rumbaugh, 1986), and this allowed us to use known and unknown lexigram stimuli in this experiment based on tests that assessed the vocabularies of these chimpanzees (Beran & Heimbauer, 2015).

Apparatus

The MTS program was written in Visual Basic 6.0. Trials were presented on a personal computer with attached 17 inch monitor. Chimpanzees responded using a digital joystick attached to their enclosure by manipulating the joystick with their hand to control a cursor on the computer screen (Figure 1, Enclosure 1). Attached to the computer was a universal feeder that dispensed food items (M&M candies) down a tube into a port in an adjacent enclosure (Figure 1, Enclosure 2). This port allowed the chimpanzees to place a hand inside to catch those food items. But if those items were not caught when dispensed, they fell into a funnel that redirected them out of the enclosure and out of the chimpanzees' reach.

Design and Procedure

At the beginning of each MTS trial, the chimpanzee was presented with a photograph in the center of the computer screen. When they touched that photograph sample using the cursor, four lexigram symbols were presented as match choices in four of six randomly chosen locations around the perimeter of the screen. These lexigrams were visuographic images that each represented a food item or object, and none of the lexigrams looked like their referents. The photograph remained visible, and the chimpanzee had to select one of the four lexigrams. The chimpanzees were very experienced in performing this task, although their knowledge of the correct lexigram for each picture varied such that some were well known and some less known.

There were two new aspects to the test that the chimpanzees had never experienced. First, after making their choice, all stimuli were erased from the screen and feedback was delayed by five seconds. After five seconds, correctly completed trials produced a melodic chime whereas incorrectly completed trials produced a buzz tone. The chimpanzees were familiar with these sounds, but had never experienced them after a delay between responding and feedback. Second, any food reward that was earned with a correct response was dispensed into an adjacent cage four seconds after the feedback sound was produced (and thus nine seconds after a matching response was made; see Figure 1). Thus, to obtain any food reward, the chimpanzees had to move to that other area before the food was dispensed. Dispensed food items fell through a tube into a port in the chimpanzee enclosure where the chimpanzee could place its hand and catch the food items. However, any items that were not caught then fell into a funnel and were redirected back out of the enclosure and onto the floor out of reach of the chimpanzee. If the chimpanzee did not catch the food when it fell, that food was forfeited and could not be obtained at any other time.

In essence, this design required that chimpanzees work in one area but be paid in another, and it required that they had to be in the correct location to obtain food when it was dispensed. So, they had to move back and forth throughout the test session, although the timing of the feedback and dispensing of food was such that they had one of two options regarding their movements. First, they could wait for the feedback to indicate whether food would be forthcoming four seconds later, and if it was, they could then quickly move to the location where it was dispensed. Four seconds was enough time for that movement if it was done briskly. Second, they could move early to the dispenser (before any feedback) and easily be in place to catch food items when dispensed; but if the response was incorrect, they would have to walk back to the test area for the next trial.

There was no penalty for incorrect matching responses, and a new trial appeared onscreen as soon as the auditory feedback was provided (i.e., even before any reward was dispensed four seconds later). Because of this, a non-confident animal could spare itself a wasted trip to the reward dispenser if an error was likely. Or at least it could wait until auditory feedback arrived to possibly justify that trip. Therefore, an optimal response pattern from the perspective of a subject that monitored its own performance would be to walk to the dispenser even before hearing any auditory feedback when it was confident a correct response had just been made. But it should wait to move at least until auditory feedback occurred on trials when it was not confident.

Each chimpanzee completed four sessions, each of 36 trials. Within these sessions, trial difficulty was determined solely by the presentation of photographs that either were associated with lexigrams known by the chimpanzees or were not. These two chimpanzees had very different rearing environments with regard to their language-acquisition projects (Rumbaugh, 1977; Savage-Rumbaugh, 1986), and also had different “vocabularies” of known lexigrams (Beran & Heimbauer, 2015). Thus, we used different photographs and lexigrams for each chimpanzee, and we included approximately half of the photographs and lexigrams that were well known by the chimpanzees and half that were not well known but that were familiar as being present on the lexigram keyboard (see Beran & Heimbauer, 2015, for extensive details on how this designation was made in formal testing). Thus, each

session involved 36 unique samples, half of which were associated with known lexigram match options and half of which were not. This allowed for the chimpanzees to experience different degrees of confidence in their responses—assuming they had some way of recognizing and assessing such confidence. Behaviorally, the animals' confidence was clearly visible given their early and late movements through space to the reward dispenser. The question was whether these movements would coordinate with their objective matching performance and their known lexigram vocabularies.

Data Scoring

In this and all experiments, a closed circuit video image of the test area was projected to a monitor in another part of the building where the experimenter observed the chimpanzees during test sessions. Thus, there was no experimenter present in the test area during these sessions¹. The experimenter watched the monitor and recorded on each trial whether the chimpanzee left the test area before or after auditory feedback. Contact with the samples also produced a spoken English word generated by the computer program for the sample (e.g., “banana”) that the experimenter recorded so as to know the sample image (which was not visible on the closed circuit camera image that the experimenter could see). These two chimpanzees have no English comprehension (Beran & Heimbauer, 2015), and although they heard the English word, it is unlikely that it played any role in their choice behavior. Because of the layout of the building (Figure 1), during real-time coding (and reliability testing), we defined a move away from the test area as the full movement of the chimpanzee's body from one enclosure into an adjacent one, and these movements were very clear.

For all experiments, an experimenter who was not present during the test sessions (BMP) reviewed all trials from one randomly selected session for each chimpanzee. She coded whether the chimpanzee moved to the dispenser before or after the auditory feedback. Reliability was very high in scoring these responses. For this experiment, the real-time scoring of early movements and the coding from the video matched on 35 of 36 trials for Lana and 36 of 36 trials for Sherman.

Results

Because there were four match choice options in each trial, chance performance was 25%. Lana and Sherman were correct on 55.6% and 75.0% of trials, respectively, in selecting the correct match choice. Thus, the procedure generated correctly completed and incorrectly completed trials for each chimpanzee. Figure 2 presents the proportion of trials on which a chimpanzee moved to the dispenser before auditory feedback for correctly completed and incorrectly completed trials. Both chimpanzees were significantly more likely to move early on trials completed correctly than on trials completed incorrectly, Lana $\chi^2(1, N = 144) =$

¹During the first test session, Sherman never left the test area even when trials were correctly completed. Thus, at the start of session two, an experimenter remained in the test area for the first five trials, and when Sherman correctly responded (and heard the tone), the experimenter urged Sherman to move to the dispensing area. This occurred on three correctly completed trials, and then the experimenter left the test area for the remainder of that and all test sessions. Note that the experimenter only helped Sherman on trials what were already correctly completed, and never encouraged Sherman to leave the test area before feedback, which was the critical measure we used to assess confidence monitoring.

58.08, $p < .001$, Sherman, $X^2(1, N = 144) = 24.00$, $p < .001$. Sherman never moved early on incorrectly completed trials; Lana rarely did so.

There was another interesting aspect of the chimpanzees' behavior. For trials in which the subjects responded correctly but did not move early, the tone should have served as an auditory cue that food was about to be delivered. On only 3 trials of 28 for Lana and 2 trials of 60 for Sherman did the chimpanzees move to the dispenser after hearing the tone indicating a correct response. For all of those trials in which they did not move, food was dispensed and was lost.

Discussion

Both chimpanzees showed the pattern expected if they were monitoring their confidence in their matching responses. They moved early, before any external feedback had been given, far more often when their response was correct than when it was incorrect. The distribution of the chimpanzees' early movements to the dispenser as a function of trial outcome was striking, and it was also striking because it was untrained. Both chimpanzees presumably learned that moving early, when appropriate, made the pace of the task less strenuous. These *confidence movements* suggest that chimpanzees know how well they have just performed in making a matching response.

An unexpected finding was that neither chimpanzee used the auditory feedback as a cue for when to move to the dispenser. The tones perfectly predicted food delivery, and the chimpanzees seemed to watch the dispenser even on trials in which they did not move, but for some reason they did not attempt to go to the dispenser after positive auditory feedback. This was despite the fact that there was time to do so, assuming they would move quickly. The chimpanzees may have found it aversive to move quickly to the dispenser after the feedback sounded, as this was not required in any other previous computerized experiment. Alternatively, these chimpanzees may not have associated strongly the feedback tones with the different outcomes in a testing context of rewards dispensed distantly.

A concern with Experiment 1 is that the chimpanzees may not have been moving on the basis of their confidence in their matching responses, but instead on the basis of past associations of certain pairings of photographs with lexigrams with strong and consistent histories of reward. In other words, the chimpanzees' early movements may not have occurred because they were confident that they had just accurately matched the correct lexigram to a photograph, but rather because they had learned that some pairings of lexigrams with photographs had produced many rewards in the past, whereas others had not. A response to this association would not require metacognition. Therefore, we designed a second experiment in which the stimuli had no past relevance for the chimpanzees and thus could not serve as cues to whether reward was likely or not. Instead, the chimpanzees had to remember stimuli during a forgetting interval within a delayed MTS test, and then possibly indicate their confidence in their own memories through their movements in space. We hypothesized that chimpanzees would continue to exhibit early movements to the food dispenser at appropriate times in the second experiment.

Experiment 2

Participants and Apparatus

Sherman and Lana again participated, along with a third chimpanzee, Mercury (27-year-old male), who had no language training (and thus could not have participated in Experiment 1). Mercury was otherwise trained to perform computerized tests such as MTS tests, including those with delays between sample presentation and the matching phase (Sherman and Lana also had similar experience with this variation). The apparatus was the same as in Experiment 1.

Design and Procedure

The task was a delayed matching-to-sample task (DMTS). At trial outset, there was a grey rectangle in the center of the screen. The chimpanzees moved the cursor into contact with that rectangle, and then it was removed to reveal a sample image on the screen (a color photograph of an object or food drawn from a library of more than 200 images). This image remained until the chimpanzees moved the cursor into contact with it. Then there followed a forgetting interval with a blank screen lasting 1, 4, 7, 10, or 13 seconds. The forgetting interval was randomly determined on each trial. We used these retention intervals to generate varying levels of matching performance by the chimpanzees, with the assumption that longer delays would produce generally lower levels of matching accuracy. After the retention interval, four choice options (all also photographs of objects or foods drawn from the same set as the sample image) appeared in four of six random locations around the perimeter of the screen, and one of those choice options was visually identical to the sample that had been presented. After the chimpanzee chose one of those four images, all images were removed from the screen, and a 5-second retention interval occurred before the presentation of auditory feedback. Finally, a 4-second retention interval occurred before the dispensing of food items (on correct trials). Each chimpanzee completed two sessions of 40 trials each. The use of such a large number of possible samples ensured that each trial was unique in terms of the sample and match choices that were presented, so that chimpanzees could not learn during the session that certain samples or matches were more or less likely to lead to food reward.

Results

For this experiment, reliability coding was conducted for one session from Sherman and one from Lana. The real-time scoring of early movements and the coding from the video matched on 40 of 40 trials for both chimpanzees.

Mercury performed at chance levels overall (27.4% correct). Unlike Sherman and Lana in Experiment 1, he never left the test area before auditory feedback, and he always went to the dispenser after hearing the tone indicating a correct response (but never after a tone indicating an incorrect response). His results are ambiguous—there is no evidence that he had remembered any of the samples let alone that he was confident he had remembered any of the samples.

The results for Sherman and Lana are shown in Figure 3. Sherman and Lana were 60.0% and 88.75% correct on matching trials, respectively. Again, chance performance was 25% because there were four match choice options in each trial. Sherman possibly showed some relation between matching performance and the length of the forgetting interval, but Lana did not—her performance was very high across all forgetting intervals. (Figure 3, top panel). Neither subject's correlation between forgetting interval and matching percentage was significant, Sherman $r(3) = -.73, p = .16$, Lana $r(3) = .79, p = .11$.

However, the crucial result was that both chimpanzees moved before auditory feedback more often on trials they had correctly answered than on trials that were incorrectly answered, Lana, $X^2(1, N = 80) = 37.58, p < .001$, Sherman $X^2(1, N = 80) = 6.88, p = .009$ (Figure 3, bottom panel). Given Lana's high performance across all retention intervals, it was not surprising that her early movements when correct also occurred at equally high levels for all retention intervals: 1 s, 3s, and 10 s intervals – 89% early movement; 7 s interval – 93% early movement; 13 s interval – 82% early movement. Sherman, however, did not move early as often as Lana, and he showed the pattern expected in which longer retention intervals also produced fewer early movements. For 1 s and 3 s retention intervals, he moved early on 58% of the trials, whereas with 7 s, 10 s and 13 s intervals, he moved early on only 31% of the trials.

Discussion

Both chimpanzees (Sherman and Lana) that learned this task moved to the dispenser early on correct trials but not on incorrect trials. As in Experiment 1, their movements reflected confidence in the memory judgment that had just been made. In this experiment, we presented trial unique comparisons of stimuli and match options, all of which were photographs, not the lexigrams that were the match options of Experiment 1. Thus, the chimpanzees had no reinforcement history from Experiment 1 in responding to those photo stimuli as match choices. Now their confidence movements could not be made on the basis of past associations to well-trained and highly rewarded stimuli (such as the lexigrams). Rather, their confidence movements had to be based on the strength of their memory for the matching choice they had just selected, or on their confidence that the matching choice would turn out to be correct.

Lana made very few errors in this experiment (88.75% correct overall) and also went early on the large majority (88.7%) of those correctly completed trials, so there were not many trials in which she needed to rely on auditory feedback that indicated she was correct. Sherman was correct overall on 60% of the trials and went early on 41.67% of correctly completed trials. There were 29 trials in which he matched correctly but did not move early (i.e., confidently). On 7 of those trials (24%), he moved after hearing the auditory feedback indicating a correct response. Thus, Sherman now showed that he would use the auditory feedback as an additional signal for moving to the dispenser, but he still relied more on his own confidence in his responses than on the external feedback.

Experiment 3

In Experiment 3, the chimpanzees again performed the DMTS task with photographs. Now, however, we varied the lengths of the forgetting intervals across individuals so as to calibrate the task difficulty to the competence level (i.e., memory abilities) of the individual chimpanzees. By finding forgetting intervals that could produce varying degrees of matching success, we hoped to produce a clearer background against which to observe the confidence movements of the chimpanzees.

We also incorporated occasional No Sample trials into the task for all chimpanzees. On these trials, the chimpanzees moved the cursor into contact with a grey rectangle that typically led to the presentation of a sample image, but now there was no sample presentation. Instead, after a 1 second delay, four match choices were presented, and one was randomly designated as the “correct” choice. Chimpanzees could not perform better than chance on these trials. Of course they should not be confident on these trials and they should not move early to the dispenser. Yet the retention interval was the shortest of any they experienced. This let us assess whether the chimpanzees would move to the dispenser based on the association formed in the Sample trials that short forgetting intervals meant frequent correct responding and reward. If so, then they should move early to the dispenser on No Sample trials. We hypothesized that the chimpanzees would move early based on confidence in their correct responses and not based on interval timing.

Participants and Apparatus

These were the same as in Experiment 2. The same large set of photographs again were used as sample and match stimuli.

Design and Procedure

The design was the same as in Experiment 2. However, now the forgetting intervals were varied across chimpanzees. For Lana, the forgetting intervals were 10, 15, 20, 25, and 30 seconds. For Sherman, they were 1, 6, 11, 16, and 21 seconds. For Mercury, they were 1, 2, 3, 4, and 5 seconds. The actual forgetting interval on each trial was determined randomly from the animal's set of forgetting intervals. No Sample trials were also presented with equal probability to Sample trials with each of the forgetting intervals. In this experiment, Lana completed 110 trials, Sherman completed 150 trials, and Mercury completed 80 trials. The varying trials numbers were the result of different degrees of effort by the chimpanzees during these scheduled test sessions (i.e., Sherman more readily engaged the task than did Lana or Mercury, and Mercury chose not to work at all in one session).

Results

Mercury still performed at chance levels for all delays (25.0% correct overall). He never left the test area before auditory feedback, but went to the dispenser after hearing the tone indicating a correct response in 100% of correct trials. For this experiment, reliability coding was conducted for one session from Sherman and one from Lana. The real-time scoring of early movements and the coding from the video matched on 40 of 40 trials for both chimpanzees.

The results for Lana and Sherman are shown in Figure 4. For Sample trials, Sherman and Lana showed a significant negative correlation of forgetting interval with percentage correct, Sherman $r(3) = -.96, p < .001$, Lana $r(3) = -.79, p < .001$ (Figure 4, top panel). Longer forgetting intervals made the matching task harder. For Sample trials, both chimpanzees again moved before hearing any auditory feedback more often on trials correctly answered than on trials incorrectly answered; Lana, $\chi^2(1, N = 96) = 4.32, p = .038$, Sherman $\chi^2(1, N = 122) = 25.46, p < .001$ (Figure 4, bottom panel). Thus, both chimpanzees again moved early more often when they likely were correct compared to when they likely were incorrect. As in Experiment 2, Lana again was equally likely to move early on correct trials for short and for long delays (range across all delays – 87% of trials to 100% of trials). Sherman, however, again showed greater likelihood of moving early on correctly completed trials with short retention intervals (1 s and 6 s – 88% of trials) than long retention intervals (16 s and 21 s – 56% of trials).

On No Sample trials, both chimpanzees performed at chance levels as expected (Sherman 25% correct, Lana 21.4% correct; Figure 4a). We compared the frequency of early moves on those trials to early moves on incorrect Sample trials. Because Lana's number of incorrect trials was low, we used a Fisher's exact test which indicated no difference in the frequency of early moves on those two types of trials, $p = .40$. There also was no difference for Sherman: $\chi^2(1, N = 60) = 0.01, p = .92$ (Figure 4b).

Discussion

Lana's and Sherman's early movements reflected confidence in the memory judgment that had just been made. Those movements were not responses simply to the shorter forgetting intervals, as chimpanzees were not any more likely to move early after the 1 s retention interval on No Sample trials than on other incorrect trials. The No Sample trials acted just like incorrect Sample trials. The 1 s retention interval on those trials gave no increased signal for moving early. To the contrary, the lack of confidence on those trials typically acted as a signal against moving early.

This was the last test given to Sherman as he clearly showed the confidence movements we had anticipated might occur. Lana and Mercury continued testing so that we could assess whether confidence movements would occur given additional task variations. For Lana, we noted an increase in movements to the dispenser on incorrect Sample trials in Experiment 3. We thought these movements might reflect her high performance levels and some inattention to her performance because she rarely made errors. Thus, we sought task parameters that would make the task more difficult for her. For Mercury, we sought task parameters that would let him show successful matching performance and possibly confidence movements as well. Here we report Lana's further results as Experiment 4, and Mercury's further results as Experiment 5.

Experiment 4

Lana was presented with black and white clip-art images drawn from a collection of hundreds of such images as samples in the DMTS task. These were meaningless images that we assumed would be more difficult to remember compared to previous experiments as they

lacked color. In her first 40-trial session, Lana was given 1, 2, 3, 4, and 5 second forgetting intervals (and also about 16% No Sample trials). She performed at 91% correct on Sample trials. The task was still too easy.

Next, we presented her with 3 sessions (a total of 120 trials) with the same stimuli but with forgetting intervals of 1, 6, 11, 16, and 21 seconds. In those three sessions, the food dispenser was in the adjacent enclosure as in all previous experiments. This is called Phase 1.

In Phase 1, we found that Lana had become highly inclined to move to the food dispenser on most trials, presumably because she was typically correct and there was not a sufficiently high cost to doing so. Thus, in subsequent sessions we presented the same task, but now the food dispenser was located two enclosures away from the computer and joystick (Figure 1, Enclosure 3). This introduced a higher cost to moving early to retrieve a potential reward. Lana completed 2 of these sessions (80 total trials). This is called Phase 2.

Results

Reliability coding was conducted for one session from Lana. The real-time scoring of early movements and the coding from the video matched on 40 of 40 trials.

For Phase 1 (Figure 5, top panel), Lana showed a significant negative correlation of forgetting interval with percentage correct for Sample trials, $r(3) = -.91, p = .031$. Now the task was more difficult at longer forgetting intervals. On Sample trials, Lana moved before auditory feedback more often on trials she had answered correctly than on trials she had answered incorrectly, $\chi^2(1, N = 104) = 11.44, p = .001$. On No Sample trials, Lana's matching performance was at chance. She did not move early more often on No Sample trials—with their 1-s retention intervals—compared to incorrectly completed Sample trials, $\chi^2(1, N = 40) = 1.37, p = .24$.

For Phase 2 (Figure 5, bottom panel), in which there was a greater cost for moving early, Lana did not show a significant negative correlation between forgetting interval and correct matching percentage, $r(3) = -.74, p = .16$, though the trend was toward longer intervals producing poorer matching performance. For Sample trials, Lana moved before hearing auditory feedback more often on trials she had correctly answered than on trials she had incorrectly answered, $\chi^2(1, N = 72) = 18.00, p < .001$. For No Sample trials, Lana necessarily performed at chance levels. She did not move early more often on No Sample trials—with their 1-s retention intervals—compared to incorrectly completed Sample trials, $\chi^2(1, N = 31) = 0.06, p = .81$. Lana's likelihood of moving early on correctly completed trials also decreased as a function of the retention interval (1 s interval – 91% of trials; 6 s interval – 93% of trials; 11 s interval – 54% of trials; 16 s interval – 63% of trials; 21 s interval – 60% of trials).

Discussion

By increasing task difficulty, and by increasing the effort and cost of confidence movements by increasing the distance to the food dispenser, we demonstrated again that Lana would monitor the confidence she had in her completed matching trials, and that she would express

that confidence through early movements to the food dispenser. Indeed, across all experiments, Lana showed this behavioral pattern that may well reflect her metacognitive monitoring of memory states and performance confidence.

Experiment 5

Design and Procedure

Mercury's testing continued with color photographs from the large set used in Experiment 2. First, we gave him one 40-trial MTS session in which the sample remained present throughout the trial. Now his matching performance improved to high levels (17 of the last 20 trials correct). Next, we gave him one DMTS session in which all trials involved a one second delay between presenting samples and match choices, and his matching performance remained high (23 of 24 trials correct), but he still always waited to hear the auditory feedback before moving to the dispenser. That is, he still made no confidence movements. So, now there was evidence that he was remembering the samples, but still no evidence that he was confident he was remembering the samples.

Next, we manipulated the interval between auditory feedback and reward, so as to put the pressure on him to move early to catch his reward before it was forfeited. We reduced this interval from four seconds to two seconds, but his pattern remained the same. Also, his strategy of moving quickly to the food dispenser after hearing the auditory feedback remained successful with the shorter delay. However, when we eliminated any delay between auditory feedback and food delivery, his pattern changed markedly. Now, Mercury had to be very close to the dispenser at the time auditory feedback was delivered if he was to catch his food reward. He could no longer rely on the auditory signal to move because it came so late it would make him late to the dispenser and thus make him miss the food reward.

Mercury completed three 0-delay sessions (60 trials each). In the first session, all Sample trials included a one second forgetting interval between sample and match-choice presentations. There were also No Sample trials as described for earlier experiments. In the second and third sessions, we introduced Sample trials involving a 3-second forgetting interval before presentation of the match choices (in addition to trials involving only a 1-second forgetting interval, as in the earlier sessions). Thus, in the entire experiment, Mercury worked on Sample trials with 1- or 3-second forgetting intervals and on No Sample trials, but he completed fewer trials with 3-second forgetting intervals than 1-sec intervals because those were only used in the last two sessions, and retention interval was randomly determined on those trials.

Results

Reliability coding was conducted for one session from Mercury. The real-time scoring of early movements and the coding from the video matched on 60 of 60 trials.

Mercury's performance is shown in Figure 6. He matched well at both 1-second and 3-second delays. He necessarily performed at chance levels on No Sample trials (13.6% correct). He was significantly more likely to move early to the dispenser when completing a

trial correctly than incorrectly with the 1-second delay, $\chi^2(1, N = 112) = 22.66, p < .001$, and with the 3-second delay, $\chi^2(1, N = 46) = 5.79, p = .016$. He did not move early more often on No Sample trials—with their 1-s retention intervals—compared to incorrectly completed Sample trials, Fisher's Exact Test, $p = .54$.

Discussion

In early testing, we found that Mercury was content to rely on external auditory signals of upcoming reward delivery when these were available and profitable. In later testing, we made the auditory signal arrive too late to allow a timely arrival at the reward dispenser, so that this feedback was no longer useable. Then, Mercury immediately came to rely on his own confidence about correct responding to signal early movement. When he was likely correct, he moved early, but when he was possibly incorrect, he did not. That is, he now made confidence movements. This change in his pattern of responding was sudden and striking. He easily made confidence movements when confidence was the only signal available or useful. This suggests that future experiments that assess confidence in nonhuman animals may require adjusting task parameters to the needs of individual animals in order to produce clear results (see also Call, 2010; Hampton et al., 2004).

General Discussion

Chimpanzees cannot provide verbal or numerical ratings of confidence in their own memories, but we gave them a way to express that confidence in the current task. They proved capable of giving confidence judgments—that is, of making confidence movements towards potential locations of food rewards—that matched their objectively-known knowledge and their present state of memory. Thus, the results indicate that chimpanzees know when they have accurately remembered and when they have not. They are able to initiate prospective reward-collecting behaviors using those psychological signals.

Experiment 1 relied on Lana's and Sherman's unique abilities to associate symbols to real-world referents. When given a matching task in which visual identity could not afford correct responding, but lexigram knowledge could (on some trials), the chimpanzees showed the confidence movements that are of central interest here. They made these movements after answering questions to which they should know the answer.

This is an important result because confidence movements were not necessary to obtain food. The auditory feedback was an objective, reliable and timely cue to signal movement to the dispenser. And yet, perhaps because this cue would have forced a faster/effortful movement, the chimpanzees largely abandoned (or did not use at all) the external signal of correctness and used the earlier available internal signal of confidence.

However, by presenting lexigram stimuli to language-trained chimpanzees, Experiment 1 offered an alternative interpretation. The chimpanzees might have been moving early (or not) based on previously associating well-known lexigrams with frequent food rewards. The subsequent DMTS tasks alleviated this concern. Sherman and Lana showed with different kinds of visual stimuli that lacked any previous reinforcement history, that their performance decreased as the memory demand increased, that their early movement was related to having

been correct on a memory trial, and that their lack of early movement was related to having been incorrect.

In contrast, Mercury initially relied on the auditory tones as signals of when to move to the dispenser. Only when we eliminated that cue's usefulness did he instead rely on an internal signal of confidence in his own responses. Then he showed confidence movements, too, ultimately providing the same evidence of memory monitoring and confidence judging.

The No Sample trials in the present studies also ruled out that chimpanzees' confidence movements were cued by the length of the retention interval in the matching trials. If they were, the chimpanzees would have moved early on those 1-s delay trials, though they could have no confidence for doing so. But they did not. We owe a debt to other researchers for knowing to include these trials. They have also been useful in memory-monitoring research with monkeys. In these studies, monkeys indicated by an uncertainty response when they do not wish to complete a memory trial (e.g., Hampton, 2001; Templer & Hampton, 2012). Monkeys treat No Sample trials in those studies as memory tests to be avoided, just as chimpanzees here treated No Sample trials as no-confidence trials. This is a strong convergence across species, and it encourages the idea that some forms of metacognition have a relatively broad distribution across the primates.

Our study is a good example of an animal-metacognitive paradigm that avoids the extensive training and associative experience that characterized many earlier paradigms (and here, we include particularly many of our own earlier studies; e.g., Smith et al., 1997, 2010). Betting responses and uncertainty responses are not highly intuitive and transparent to animals, and their training is difficult and often extensive. And with the training could come the unwanted consequences of associative cues and responses. In contrast, the present task required a movement through space to operate as the behavioral manifestation of confidence, which made good use of the natural tendencies of these animals to move towards potential locations of food. Wild chimpanzees travel widely for foraging, and presumably begin movements over long distances on the basis of remembering where certain foods are located, and perhaps even whether such foods are at a peak point of ripeness or growth (Ban, Boesch, & Janmaat, 2014). Moving to food that is expected to be there when one arrives, and then having that food there on arrival would serve a clear adaptive purpose.

Thus, the naturalistic responses of interest in terms of movements towards food revealed metacognitive phenomena with almost no training. Even in Mercury's case, the metacognitive performance emerged immediately when the circumstances required it. Here, too, we owe a debt to other researchers. Call and his colleagues transcended training-intensive paradigms by studying chimpanzees' searches for food hidden in opaque tubes. They demonstrated uncertainty monitoring by chimpanzees with almost no training (Call & Carpenter, 2001). However, that paradigm faced a criticism that perhaps the apes employed a generalized search strategy that led to responses focused on looking for food rather than responses that emerged because of a metacognitive monitoring strategy (e.g., Carruthers, 2008; Crystal & Foote, 2011). This alternative account has been assessed subsequent to the early reports (Call, 2010; Marsh & MacDonald, 2012a), but the present study is immune in its design to that particular criticism because the chimpanzees in this study could not use any

generalized strategy to look for food (see also Beran et al., 2013). And, the present study allowed for chimpanzees to show, in a qualitatively different kind of test, a new, untrained response pattern (confidence movements) that one would also expect if metacognitive monitoring were occurring.

Claims of metacognitive abilities in animals face the appropriate burden of needing to result from a variety of testing paradigms, and response outputs, that allow one to argue that the preponderance of the evidence supports such a conclusion. This research area has faced some criticism that often appeals to Morgan's Canon for constraining the interpretations of researchers in this area. However, Morgan's Canon allows for interpretations at more complex levels when the data consistently support such interpretations across a wide range of independent tests (see Smith, Couchman, & Beran, 2012; Sober, 1998). For this reason, the present results, in relation to those from other tests with apes in different kinds of metacognition tests, and those with monkeys and other animals in metamemory tests and perceptual discrimination tasks, converge on the idea that animals share with humans the capacity for metacognitive monitoring. Such a capacity need not manifest equally across species in terms of the proficiency with which it is applied to problems or in terms of the conscious experiences that may or may not accompany its manifestation.

We point out that the need for intuitive paradigms that require little to no training is likely to be especially crucial in the study of chimpanzees' metacognition. From our experience, our chimpanzees simply will not do trial-intensive psychophysical tasks such as those that we have used with macaques. We also point out that intuitive paradigms may be especially crucial in the domain of animal metacognition itself, because many metacognitive responses may be among the most difficult, cognitively effortful, and working-memory intensive responses that animals can make. These responses can be especially elusive in some tasks, needing particular empirical fostering and situational support. In fact, this elusiveness may itself have theoretical implications about the cognitive sophistication of those responses when they occur, and about their cognitive level and organization. More broadly, using situations and tasks that allow animals to move to food when they "should" may open new avenues for investigating what animals know about their own perceptual and conceptual abilities and memories. Perhaps other species also will show that they will *go when they know*, and such outcomes will help define the breadth and depth of comparative metacognition.

Chimpanzees' confidence movements highlight their ability to monitor their knowledge, and our results complement others showing that chimpanzees know when they need information (Beran et al., 2013; Call, 2010; Call & Carpenter, 2001). The consensus is growing that animals can show a degree of metacognitive monitoring. However, this consensus—even if animals make confidence judgments, seek needed information, and reflect on what they do or do not know—does not imply that animal metacognition is synonymous with human metacognition. It is important to say that these findings do not reveal the experiential aspects of animal metacognition. Animals like chimpanzees may or may not be fully conscious of being confident. They may not have self-awareness as they judge confidence. They may not have meta-representation as they process their metacognitive states. Human metacognition has many facets—animals could share some but not all of these components. The

preliminary evidence is that some nonhuman primates share with humans an uncertainty-confidence cognitive resource. It is an exciting time in the developing field of animal metacognition, as researchers now ask whether research of this kind can extend further to ask more direct questions about the degree to which animals may have conscious, declarative, and self-reflective metacognition.

Acknowledgments

This research was supported by NIH grants HD061455 and HD060563 and by NSF grant BCS0956993.

References

- Adams A, Santi A. Pigeons exhibit higher accuracy for chosen memory tests than for forced memory tests in duration matching-to-sample. *Learning and Behavior*. 2011; 39:1–11. [PubMed: 21264568]
- Balcomb FK, Gerken L. Three-year-old children can access their own memory to guide responses on a visual matching task. *Developmental Science*. 2008; 11:750–760. [PubMed: 18801131]
- Ban SD, Boesch C, Janmaat KRL. Tai chimpanzees anticipate revisiting high-value fruit trees from further distances. *Animal Cognition*. 2014 published online. 10.1007/s10071-014-0771-y
- Basile BM, Hampton RR, Suomi SJ, Murray EA. An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition*. 2009; 12:169–180. [PubMed: 18712532]
- Benjamin AS, Bjork RA, Schwartz BL. The mismeasure of memory: When retrieval fluency is misleading as a metacognitive index. *Journal of Experimental Psychology: General*. 1998; 127:55–68. [PubMed: 9503651]
- Beran MJ. Quantity perception by adult humans (*Homo sapiens*), chimpanzees (*Pan troglodytes*), and rhesus macaques (*Macaca mulatta*) as a function of stimulus organization. *International Journal of Comparative Psychology*. 2006; 19:386–197.
- Beran MJ, Heimbauer LA. A longitudinal assessment of vocabulary retention in symbol-competent chimpanzees (*Pan troglodytes*). *PLoS ONE*. 2015; 10:e0118408. [PubMed: 25706561]
- Beran MJ, Pate JL, Washburn DA, Rumbaugh DM. Sequential responding and planning in chimpanzees (*Pan troglodytes*) and rhesus macaques (*Macaca mulatta*). *Journal of Experimental Psychology: Animal Behavior Processes*. 2004; 30:203–212. [PubMed: 15279511]
- Beran MJ, Rumbaugh DM. “Constructive” enumeration by chimpanzees (*Pan troglodytes*) on a computerized task. *Animal Cognition*. 2001; 4:81–89.
- Beran MJ, Smith JD. Information seeking by rhesus monkeys (*Macaca mulatta*) and capuchin monkeys (*Cebus apella*). *Cognition*. 2011; 120:90–105. [PubMed: 21459372]
- Beran MJ, Smith JD, Perdue BM. Language-trained chimpanzees name what they have seen, but look first at what they have not seen. *Psychological Science*. 2013; 24:660–666. [PubMed: 23508741]
- Beran MJ, Smith JD, Redford JS, Washburn DA. Rhesus macaques (*Macaca mulatta*) monitor uncertainty during numerosity judgments. *Journal of Experimental Psychology: Animal Behavior Processes*. 2006; 32:111–119. [PubMed: 16634654]
- Beran MJ, Washburn DA. Chimpanzee responding during matching-to-sample: Control by exclusion. *Journal of the Experimental Analysis of Behavior*. 2002; 78:497–508. [PubMed: 12507016]
- Bjork RA, Dunlosky J, Kornell N. Self-regulated learning: Beliefs, techniques, and illusions. *Annual Review of Psychology*. 2012; 64:417–444.
- Call J. Do apes know that they could be wrong? *Animal Cognition*. 2010; 13:689–700. [PubMed: 20306102]
- Call J, Carpenter M. Do apes and children know what they have seen? *Animal Cognition*. 2001; 4:207–220.
- Carruthers P. Meta-cognition in animals: A skeptical look. *Mind and Language*. 2008; 23:58–89.
- Carruthers P. How we know our own minds: The relationship between mindreading and metacognition. *Behavioral and Brain Sciences*. 2009; 32:121–182. [PubMed: 19386144]

- Castro L, Wasserman EA. Information-seeking behavior: Exploring metacognitive control in pigeons. *Animal Cognition*. 2013; 16:241–254. [PubMed: 23065186]
- Crystal JD. Where is the skepticism in animal metacognition? *Journal of Comparative Psychology*. 2014; 128:152–154. [PubMed: 24866006]
- Crystal JD, Foote AL. Metacognition in animals: Trends and challenges. *Comparative Cognition and Behavior Reviews*. 2009; 4:54–55.
- Crystal JD, Foote AL. Evaluating information-seeking approaches to metacognition. *Current Zoology*. 2011; 57:531–542.
- Dunlosky, J.; Bjork, RA., editors. *Handbook of memory and metamemory*. New York: Psychology Press; 2008.
- Flavell JH. Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*. 1979; 34:906–911.
- Foote AL, Crystal JD. Metacognition in the rat. *Current Biology*. 2007; 17:551–555. [PubMed: 17346969]
- Foote AL, Crystal JD. “Play it again”: A new method for testing metacognition in animals. *Animal Cognition*. 2012; 15:187–199. [PubMed: 21909935]
- Fujita K. Metamemory in tufted capuchin monkeys (*Cebus apella*). *Animal Cognition*. 2009; 12:575–585. [PubMed: 19242741]
- Hampton RR. Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*. 2001; 98:5359–5362.
- Hampton RR. Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms? *Comparative Cognition and Behavior Reviews*. 2009; 4:17–28. [PubMed: 20046911]
- Hampton RR, Hampstead BM. Spontaneous behavior of a rhesus monkey (*Macaca mulatta*) during memory tests suggests memory awareness. *Behavioural Processes*. 2006; 72:184–189. [PubMed: 16460887]
- Hampton RR, Zivin A, Murray EA. Rhesus monkeys (*Macaca mulatta*) discriminate between knowing and not knowing and collect information as needed before acting. *Animal Cognition*. 2004; 7:239–246. [PubMed: 15105996]
- Inman A, Shettleworth SJ. Detecting metamemory in nonverbal subjects: A test with pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*. 1999; 25:389–395.
- Iwasaki S, Watanabe S, Fujita K. Do pigeons (*Columba livia*) seek information when they have insufficient knowledge? *Animal Cognition*. 2013; 13:211–221. [PubMed: 23065184]
- Jozefowicz J, Staddon JER, Cerutti DT. Metacognition in animals: How do we know that they know? *Comparative Cognition and Behavior Reviews*. 2009; 4:29–39.
- Kirk CR, McMillan N, Roberts WA. Rats respond for information: Metacognition in a rodent? *Journal of Experimental Psychology: Animal Learning and Cognition*. 2014; 40:249–259. [PubMed: 24893220]
- Koriat A, Goldsmith M. Memory in naturalistic and laboratory contexts: distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*. 1994; 123:297–315. [PubMed: 7931094]
- Koriat A, Goldsmith M. Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*. 1996; 103:490–517. [PubMed: 8759045]
- Kornell N. Metacognition in humans and animals. *Current Directions in Psychological Science*. 2009; 18:11–15.
- Kornell N. Where is the “meta” in animal metacognition? *Journal of Comparative Psychology*. 2014; 128:143–149. [PubMed: 23875919]
- Kornell N, Son LK. Learners’ choices and beliefs about self-testing. *Memory*. 2009; 17:493–501. [PubMed: 19468957]
- Kornell N, Son LK, Terrace HS. Transfer of metacognitive skills and hint seeking in monkeys. *Psychological Science*. 2007; 18:64–71. [PubMed: 17362380]

- Le Pelley ME. Metacognitive monkeys or associative animals? Simple reinforcement learning explains uncertainty in nonhuman animals. *Journal of Experimental Psychology, Learning, Memory, and Cognition*. 2012; 38:686–708.
- Marsh HL, MacDonald SE. Information seeking by orangutans: A generalized search strategy? *Animal Cognition*. 2012a; 15:293–304. [PubMed: 21918872]
- Marsh HL, MacDonald SE. Orangutans (*Pongo abelii*) “play the odds”: Information-seeking strategies in relation to cost, risk, and benefit. *Journal of Comparative Psychology*. 2012b; 126:263–278. [PubMed: 22023265]
- Metcalfe, J.; Kober, H. Self-reflective consciousness and the projectable self. In: Terrace, HS.; Metcalfe, J., editors. *The missing link in cognition: Origins of self-reflective consciousness*. New York, NY: Oxford University Press; 2005.
- Morgan G, Kornell N, Kornblum T, Terrace HS. Retrospective and prospective metacognitive judgments in rhesus macaques (*Macaca mulatta*). *Animal Cognition*. 2014; 17:249–257. [PubMed: 23812677]
- Nakamura N, Watanabe S, Betsuyaku T, Fujita K. Do birds (pigeons and bantams) know how confident they are of their perceptual decisions? *Animal Behaviour*. 2011; 14:83–93.
- Nelson, TO., editor. *Metacognition: Core readings*. Toronto: Allyn and Bacon; 1992.
- Nelson TO, Dunlosky J. When people’s judgments of learning (JOL) are extremely accurate at predicting subsequent recall: The delayed-JOL effect. *Psychological Science*. 1991; 2:267–270.
- Reder LM. Strategy selection in question answering. *Cognitive Psychology*. 1987; 19:90–138.
- Roberts WA, Feeney MC, McMillan N, MacPherson K, Musolino E, Petter M. Do pigeons (*Columba livia*) study for a test? *Journal of Experimental Psychology: Animal Behavior Processes*. 2009; 35:129–142. [PubMed: 19364222]
- Rumbaugh, DM. *Language learning by a chimpanzee: The LANA Project*. New York: Academic Press; 1977.
- Savage-Rumbaugh, ES. *Ape language: From conditioned response to symbol*. New York: Columbia University Press; 1986.
- Schwartz BL. Working memory load differentially affects tip-of-the-tongue states and feeling-of-knowing judgments. *Memory & Cognition*. 2008; 36:9–19. [PubMed: 18323058]
- Shields WE, Smith JD, Guttmanova K, Washburn DA. Confidence judgments by humans and rhesus monkeys. *Journal of General Psychology*. 2005; 132:165–186. [PubMed: 15871299]
- Shields WE, Smith JD, Washburn DA. Uncertain response by humans and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task. *Journal of Experimental Psychology: General*. 1997; 126:147–164. [PubMed: 9163934]
- Smith JD. The study of animal metacognition. *Trends in Cognitive Sciences*. 2009; 13:389–396. [PubMed: 19726218]
- Smith JD, Beran MJ, Couchman JJ, Coutinho MVC. The comparative study of metacognition: Sharper paradigms, safer inferences. *Psychonomic Bulletin and Review*. 2008; 15:679–691. [PubMed: 18792496]
- Smith JD, Beran MJ, Redford JS, Washburn DA. Dissociating uncertainty responses and reinforcement signals in the comparative study of uncertainty monitoring. *Journal of Experimental Psychology: General*. 2006; 135:282–297. [PubMed: 16719654]
- Smith JD, Couchman JJ, Beran MJ. The highs and lows of theoretical interpretation in animal-metacognition research. *Philosophical Transactions of the Royal Society B*. 2012; 367:1297–1309.
- Smith JD, Redford JS, Beran MJ, Washburn DA. Rhesus monkeys (*Macaca mulatta*) adaptively monitor uncertainty while multi-tasking. *Animal Cognition*. 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*). *Journal of Experimental Psychology: General*. 1995; 124:391–408. [PubMed: 8530911]
- Smith JD, Shields WE, Allendoerfer KR, Washburn DA. Memory monitoring by animals and humans. *Journal of Experimental Psychology: General*. 1998; 127:227–250. [PubMed: 9742715]

- Smith JD, Shields WE, Schull J, Washburn DA. The uncertain response in humans and animals. *Cognition*. 1997; 62:75–97. [PubMed: 8997171]
- Sober, E. Morgan's canon. In: Cummins, D.; Allen, C., editors. *The evolution of mind*. New York, NY: Oxford University Press; 1998. p. 224–242.
- Suda-King C. Do orangutans (*Pongo pygmaeus*) know when they do not remember? *Animal Cognition*. 2008; 11:21–42. [PubMed: 17437141]
- Suda-King C, Bania AE, Stromberg EE, Subiaul F. Gorillas' use of the escape response in object choice memory tests. *Animal Cognition*. 2013; 16:65–84. [PubMed: 22923213]
- Sutton JE, Shettleworth SJ. Memory without awareness: Pigeons do not show metamemory in delayed matching to sample. *Journal of Experimental Psychology: Animal Behavior Processes*. 2008; 34:266–282. [PubMed: 18426309]
- Templer VL, Hampton RR. Rhesus monkeys (*Macaca mulatta*) show robust evidence for memory awareness across multiple generalization tests. *Animal Cognition*. 2012; 15:409–419. [PubMed: 22183204]

Highlights

- Humans can take actions that manifest belief in likely success, and this reflects metacognition.
- Chimpanzees performed computerized tasks and had to move to another location to receive food rewards for correct responses.
- The chimpanzees moved to the food delivery location more often on correctly completed trials than incorrectly completed trials, even before the computer provided any feedback about the accuracy of the response.
- Chimpanzees monitored their own performance and spontaneously expressed their confidence in their responses by moving – or not – to the location of food delivery.

Figure 1.

A schematic of the testing area within the larger chimpanzee housing area. Sections shown in grey housed the chimpanzees, and three enclosures were used, each of which connected to the adjacent enclosure by a doorway through which the chimpanzees could move. TV shows the area in which the experimenters observed ongoing test sessions via a closed-circuit monitor, and this area was behind an opaque wall that kept chimpanzees from seeing them (the dark vertical line). J represents where the joystick and computer apparatus were setup, and D represents the two locations where the dispenser could be located depending on how far the chimpanzees had to move to retrieve food rewards. C represents the location of the camera that recorded the chimpanzees.

Figure 2.

The performance of Lana and Sherman in Experiment 1 in terms of the percentage of trials in which they moved early to the food dispenser, shown for correctly completed trials and incorrectly completed trials.

Figure 3.

The performance of Lana and Sherman in Experiment 2. The top panel shows the percentage of correct matching choices as a function of the retention interval between presentation of the sample and match choices. The bottom panel shows the percentage of trials in which they moved early to the food dispenser, shown for correctly completed trials and incorrectly completed trials.

Figure 4.

The performance of Lana and Sherman in Experiment 3. The top panel shows the percentage of correct matching choices as a function of the retention interval between presentation of the sample and match choices, including for trials with no sample. The bottom panel shows the percentage of trials in which they moved early, shown for correctly completed trials and incorrectly completed trials with samples, and for trials without a sample.

Figure 5.

The performance of Lana in Experiment 4. The top panel shows Lana's percentage of correct matching choices as a function of the retention interval between presentation of the sample and match choices for sessions where she had to move only to an adjacent cage or two cages away. The bottom panel shows the percentage of trials in which she moved early, shown for correctly completed trials and incorrectly completed trials with samples, and for trials without a sample, and as a function of how far she had to move.

Figure 6.

The performance of Mercury in Experiment 5. The top panel shows the percentage of correct matching choices as a function of the retention interval between presentation of the sample and match choices, including for trials with no sample. The bottom panel shows the percentage of trials in which Mercury moved early, shown for correctly completed trials and incorrectly completed trials with samples, and for trials without a sample.