2010

Uncertainty Monitoring in Sprague-Dawley Rats (Rattus norvegicus)

Leslie Anne Angel
The University of Montana

UNCERTAINTY MONITORING IN SPRAGUE-DAWLEY RATS (*RATTUS*

NORVEGICUS)

By

Leslie A. Angel

M.A., University of Dayton, Dayton, Ohio, 2003
B.A., Wright State University, Fairborn, Ohio, 2001

Dissertation

presented in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy
in Experimental Psychology

The University of Montana
Missoula, MT

August 2010

Approved by:

Perry Brown, Associate Provost for Graduate Education
Graduate School

Dr. Wendy Shields, Chair
Psychology

Dr. Allen Szalda-Petree
Psychology

Dr. Nabil Haddad
Psychology

Dr. Stuart Hall
Psychology

Dr. Irene Appelbaum
Anthropology

Angel, Leslie A., Ph.D., August 2010   Experimental Psychology

Uncertainty Monitoring in Sprague-Dawley Rats *(Rattus norvegicus).*

Chairperson: Dr. Wendy Shields

Comparative psychologists have explored the metacognitive capabilities of rhesus monkeys, a capuchin monkey, an orangutan, a dolphin, pigeons, and rats. Previous research with rats has demonstrated inconsistent results (Foote & Crystal, 2007; Smith & Schull, 1989). In the current study, two Sprague-Dawley rats were tested in a tone-length discrimination task in which they were prompted to press one of two levers indicating "short" or "long" depending on the duration of a tone. They also had the option to opt out of some trials. If the subjects knew when they did not know the answer to the task, they were expected to opt-out more frequently as the difficulty of the task increased. They were also expected to demonstrate differences in accuracy for trials during which they could opt-out and trials in which they could not opt-out. Higher accuracies were expected on trials during which they could opt-out. One subject did not select the opt-out option during testing. The other subject did not use the opt-out option adaptively by opting out more as the difficulty of the stimulus discrimination increased. However, when comparing trials in which the subjects could not opt-out and those in which the subjects could opt-out, this subject demonstrated higher accuracies on trials in which he could opt-out. This provides some evidence that at least one rat knows when he does not know the answer to a duration-discrimination task. This experiment imparts clarification to previous research and provides further evidence for uncertainty monitoring among rats, lending greater understanding to the evolutionary development of uncertainty monitoring.

DEDICATION

This manuscript is dedicated to Earl ("Pop") Angel and Eric Ramsey. Your love of science, love of life, and unconditional love for me guided me to where I am today and will always be with me. I miss you both each day.

"Here's to the hearts and the hands of the men, that come with the dust and are gone with the wind." –Bob Dylan's *Song to Woody*

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

Uncertainty Monitoring in Sprague-Dawley Rats *(Rattus norvegicus)*

Comparative research in metacognition has provided compelling insights into the phylogeny of at least one aspect of metacognitive processing (see Smith, Shields, and Washburn, 2003). Specifically, however, previous research with rats has revealed perplexing and contradictory results. The aim of the current study is to provide clarification of these results and further insight into uncertainty monitoring among rats.

What is Metacognition?

J. H. Flavell first formally used the term metacognition in 1976 to refer to knowledge of one's own cognitive processes. Metacognition is casually referred to as cognition about cognition, or thinking about thinking, and can also be defined as knowledge and control of one's own cognitive processing (Brown, 1987; Flavell, 1979; Nelson, Narens, & Dunlosky, 2004). Flavell (1979) argues that there are two key aspects of metacognition: metacognitive knowledge and metacognitive experience. Metacognitive knowledge involves one's understanding of his or her overall knowledge of cognitive or psychological phenomena (Flavell, 1987). An example of this is an adult who recognizes that a specific memory may be fallible because she recalls previous errors in her own memory. Metacognitive experiences refer rather to current, conscious processing. This processing could be cognitive or affective. One example could be a student who recognizes that the content of the lecture she is listening to is becoming increasingly more difficult. Flavell maintains that both metacognitive knowledge and experience are key to metacognition.

Several important distinctions need to be made about metacognitive processes. First, metacognition refers specifically to knowledge and experience of one's own

cognitive states rather than knowledge of others' cognitive states. The latter is typically referred to as "theory of mind" and is thought to involve distinct processing from metacognition (e.g., Carruthers, 2008; Flavell, 1976; Premack & Woodruff, 1978). Secondly, metacognitive processes are also distinct from cognitive processes, although overlap certainly occurs (Brown, 1987). For example, metacognition is often described as involving processes such as attention, conflict resolution, error correction, inhibitory control, and emotional regulation, all processes that may also be involved in various cognitive processes (see Shimamura, 2000). Metacognition is distinct in that it involves thought about one's own thinking, or a self-reflective quality. For example, a student thinking about last night's biological psychology lecture demonstrates a cognitive process, but his awareness that he can't recall the term "long-term potentiation" even though he remembers hearing the term in class and knows that he knows it, is a metacognitive judgment. What is metacognitive is cognitive, but what is cognitive is not always metacognitive.

Metacognition is often used to discuss awareness involved in knowing what one knows and knowing what one does not know (see Smith, Shields, & Washburn, 2003). Feeling of knowing (FOK) is characterized by feelings of certainty about information stored in memory, such as the name of an acquaintance, even as one is unable to recall the information (Hart, 1965; Koriat, 1993, 1994, 1995; Miner & Reder, 1994; Nelson & Narens, 1990; Reder, 1994). During a FOK, the individual is unable to retrieve information, although a feeling that the information resides in memory and is only temporarily unattainable persists. One may be able to recall specifics about the intended memory, such as a first letter of the sought after name or the context associated with the

2

name (see Hart, 1965). Indeed, one specific aspect of FOK is the tip-of-the-tongue (TOT) phenomenon. William James (1890) was one of the first to write about the familiar TOT experience in his text, *The Principles of Psychology*.

Evolutionary Roots of Metacognition

Metacognition has been used as a distinguishing factor between humans and nonhuman animals, similar to the classic examples of tool use and language and is often touted as one of the hallmarks and highest levels of human information processing (see Carruthers, 2008). It is reasonable to theorize that metacognitive processes and mechanisms may have also evolved along a continuum, such that evidence for metacognitive capabilities may exist in other species. Considerable effort has been made in recent years to find a neurological basis for metacognition (Chua, Schacter, & Sperling, 2009; Fernandez-Duque, Baird, & Posner, 2000; Gallo, Kensinger & Schacter, 2006; Janowsky, Shimamura, & Squire, 1989; Kepecs, Uchida, Zariwala, & Mainen, 2008; Shimamura, 2000). Neurological evidence revealing similarities between humans and nonhumans has been discovered in areas associated with conscious processing, including the thalamic midline and intralaminar nuclei (Seth, Edelman, & Baars, 2004; Shimamura, 1994; Van der Werf et al., 2002). Further discovery of similarities and differences involved in metacognitive processes in human and nonhuman animals would lend greater understanding to the evolutionary development of metacognition and could reveal compelling insights into the phylogeny of metacognitive processes.

Metacognitive capabilities, such as suspending responses, requesting additional information, and indicating certainty or uncertainty appear to be adaptive functions for human adults and older children (e.g., McGuinness, 2005; Smith, Shields, & Washburn,

2003). Because metacognition was traditionally categorized as an exclusively human

capability, indicative of higher-level thinking, an important next step in this line of

research was to determine whether other species would display metacognitive processing

capabilities.

Comparative Research in Uncertainty Monitoring

Much of the history of metacognitive research in humans has measured

metacognition through procedures that require language, such as self-assessment

questionnaires, pre- or post-experimental questions, and thinking-aloud measurements

(e.g., Kreutzer, Leonard, & Flavell, 1977; Markman, 1979; McGuinness, 1990; Nelson &

Narens, 1990; Wellman, 1977). Humans are unique in their ability to report when they

know something, as well as to report when they do not know (see Smith, Shields, &

Washburn, 2003). A verbal description of not knowing is often considered to be

evidence of subjective uncertainty. For example, when faced with a challenging trivia

question, such as "Which element from the periodic table makes up diamonds?" if one is

unsure of the answer, then one is typically well aware of his or her subjective uncertainty.

This experience often is accompanied by affective consequences, such as frustration.

Reporting both the cognitive (e.g., inability to access stored information) and affective

(e.g., frustration) aspects of subjective uncertainty is often provided as evidence of

metacognitive processes in humans.

However, because language can be misinterpreted, misrepresented, biased, and

even reproduced in non-living systems such as computers, further evidence of

metacognition should be considered. Even though language has been used as a

measurement of metacognition in humans, lacking language skills does not necessarily

mean that an organism lacks metacognition or at least some aspects of metacognitive processing (see Smith, Shields, & Washburn, 2003). Therefore, there was a need to develop tasks capable of measuring metacognition among nonverbal species.

Comparative research in metacognitive judgments has revealed considerable evidence for at least some aspects of metacognitive processing in nonhuman animals (see Smith, Shields, & Washburn, 2003, but see also Carruthers, 2008). For example, metacognitive research employing perceptual, problem solving tasks have revealed similar patterns of responding in rhesus monkeys, a capuchin, an orangutan, a dolphin, and humans (e.g., Beran, Smith, Coutinho, Couchman, & Boomer, 2009; Smith, 2009; Smith, Schull, Strote, McGee, Egnor, & Erb, 1995; Smith, Shields, Schull, & Washburn, 1997; Suda-King, 2008; Washburn et al., 2006). The literature also extends to other mammalian and non-mammalian species in search of evolutionary roots for metacognitive ability. These studies broaden metacognitive research to pigeons and rats (e.g., Inman & Shettleworth, 1999; Sole, Shettleworth, & Bennett, 2003; Smith & Schull, 1989; Foote & Crystal, 2007). As will be discussed in more detail, the limited research involving metacognitive processing in rats is fraught with inconsistencies.

Nonverbal tasks have been used to test metacognitive processing in both humans and nonhumans (e.g., Hampton, 2001; Shields, 1999; Shields, Smith, & Washburn, 1997; Smith & Schull, 1989; Smith, Shields, Allendoerfer, & Washburn, 1998; Smith et al., 1995; Sole, Shettleworth, & Bennett, 2003; Washburn et al., 2006). Nonverbal tasks used to test for metacognition involve two important steps, the presentation of a problem solving task that creates ambiguity and the option to decline (hereafter, opt out) or take a test (hereafter, opt in; Shields, Smith, & Washburn, 1997; Smith, Shields, & Washburn,

2003). The problem solving task must include different levels of difficulty. Variation in difficulty is necessary to measure distinctions in performance between opt-in and opt-out trials. If the participants know that they know the answer, they should opt in, whereas, if they know that they do not know the answer, they should opt out. In these experiments, performance was expected to decline for more difficult trials and opting out was expected to increase as difficulty of trials increases.

Nonverbal metacognition experiments can vary in a number of ways, such as the type of task employed (memory-based or perceptual) and the timing of the prompt for an answer or opt-out option relative to the question (see Smith, Shields, & Washburn, 2003). It is the timing of the prompt for an answer that is most relevant to the following line of research. One approach is to provide subjects the opportunity to opt in or opt out at the same time the answers are available. Smith, Shields, Allendoerfer, and Washburn (1998) employed this task-type to test uncertainty monitoring among rhesus monkeys. This experiment tested two rhesus monkeys in a memory-based, serial probe recognition (SPR) task in which the prompt for an answer and the opt-out response were posed simultaneously. In an SPR task, a series of items to be remembered is presented. Then, a probe item is displayed and subjects are prompted to indicate whether the probe was included in the last set or not. In this case, pictures of polygons were first presented. Then, a probe, a polygon that had either been presented in the last set or not was displayed. Monkeys (as well as humans) often reveal a pronounced primacy and recency effect in SPR tasks. When graphed, this creates a characteristic "U-shaped" curve in which the middle items in a list are recalled less frequently than the first and last items of

the list. In this experiment, an opt-out option allowed the monkeys to "escape" into a

guaranteed win trial.

Results revealed a U-shaped serial position curve and an opt-out curve that

mirrored it. As accuracy in memory declined for the middle items in the list, the opt-out

rate increased. Opting out occurred less frequently for the most recalled items of the list

(i.e., the earliest and most recent items). The monkeys used the opt-out option in an

adaptive way, to avoid trials in which difficult serial positions were probed and during

which memory was relatively poor. These results were compared to human performance.

Human performance was similar to that of the rhesus monkeys in uncertainty monitoring

when employing a serial probe recognition.

In another simultaneous judgment-and-test design, Smith, Shields, Schull, and

Washburn (1997) employed a psychophysical task to test uncertainty monitoring in

rhesus monkeys. In this study, two rhesus monkeys were tested in a psychophysical

density-discrimination task. They chose one response option if exactly 2,950 pixels were

illuminated in a box on a screen or another option if fewer than 2,950 pixels were

illuminated. Difficulty of the task was based on titration of pixels, such that 2,000 pixels

(further from threshold) would be easier to identify than 2,500 pixels (closer to

threshold). Thus, the range of stimuli prone to error could be empirically determined. A

third, opt-out option led to an easy, guaranteed win trial. The monkeys declined trials

most frequently at the perceptual threshold. Thus, they used the opt-out response

adaptively on these difficult trials. Results were compared to human performance in a

similar task. Both monkeys and humans were found to use the opt-out option for trials

that fell in the uncertain range of stimuli.

Much of the research in metacognition among humans involves "prospective tasks" (see Nelson & Narens, 1990). Prospective procedures involve asking subjects a general question first, then presenting subjects with an option to opt in or out, and finally cueing subjects for an answer (e.g., Hart, 1965; Nelson & Narens, 1990). Hence, prospective tasks require subjects to make judgments about their knowledge before presenting them with answer alternatives. In this way, prospective tasks are fairly complex because subjects must determine their level of confidence prior to the presentation of the potential answers.

Several studies used a prospective task to test memory monitoring among rhesus monkeys (Hampton, 2001; Shields, 1999). Using a paired-associate task with polygon shapes, Shields (1999) tested a monkey's ability to link a sample polygon to a target. The correct match to the target had to be learned as it was based on correct associations between the target and the sample on previous trials. This task requires memory for previous matches. Some associations were displayed less frequently than others, increasing their difficulty. A sample polygon (opt in) and a star (opt out) were displayed on a screen (judgment phase). Moving a cursor to the sample polygon indicated a judgment of knowing (opt in) and led to a second screen (the test phase) with the correct match to the sample (the target) and an incorrect foil. Moving the cursor to the star during the judgment phase indicated a judgment of not knowing (opt out), and also led to the test phase. Correct responses led to a food reward in both the opt-in trials and the opt-out trials. Incorrect responses during the opt-in trials led to a long timeout whereas incorrect responses during opt-out trials led to a very brief timeout.

Results indicated that the monkey's accuracy decreased on more difficult trials for both the opt-in and opt-out trials. However, if the monkey was making decisions based on metacognition, one would expect to see more opt-outs during trials involving associations that were presented less frequently and were, therefore, more difficult. Yet, the opt-out rate was not related to the trial difficulty. The monkey was not declining more difficult trials and did not use the opt-out response as a judgment-of-not-knowing. Further, the monkey's accuracy for trials taken and trials escaped were nearly identical. Accuracy comparisons between opt-in and opt-out trials are crucial in metacognition experiments. These accuracy comparisons determine whether the opt-out option is being used adaptively because one would expect performance to be better for opt-in trials and worse for opt-out trials if participants know what they know and know what they do not know. Humans tested in a similar task showed greater accuracy for opt-in trials than opt-out trials, as would be expected if one were to use the opt-out response as a tool to deal with difficult, error-prone trials. The monkey was not helping himself by opting out, but rather seemed to be using the opt-out response arbitrarily across trials. The monkey was not using the opt-out response adaptively as do most humans in tasks involving prospective memory monitoring.

Hampton (2001) conducted a similar study testing memory monitoring in two rhesus monkeys using a delayed matching-to-sample (DMTS) task. A sample was displayed on a screen during a study phase and was removed to create a forgetting interval. The forgetting interval varied from 15 s to over 100 s. After the forgetting interval, the monkeys were given a choice between taking (opting in) or declining (opting out of) the test in two-thirds of the trials (Choice trials). Taking the test resulted in a

desired food reward for correct matches or a timeout for incorrect matches. Declining the memory test resulted in a less desired food reward and no risk of a timeout. In the other one-third of the trials, the monkeys had to take the test (Forced trials).

The monkeys were expected to decline more difficult choice trials, or those trials in which there was a longer forgetting interval (Hampton, 2001). Also, similar to the accuracy comparisons made between taken and declined trials in Shields' (1999) study, Choice trials were predicted to produce better performance than Forced trials because monkeys could decline trials for which they did not know the correct match to the sample. One monkey performed consistently with both the predictions based on metamemory. He declined difficult trials more frequently, and his accuracy for trials in which he could not escape was lower than that of trials in which he could opt out. The other monkey declined more difficult trials, but did not perform better on Choice trials compared to Forced trials.

A third and final approach has been used to test uncertainty by changing the order in which answers and judgments are prompted following a question. A retrospective task is one in which a primary discrimination response (e.g., categorizing stimuli as dense or sparse) is followed by a confidence rating (Shields, Smith, Guttmannova, and Washburn, 2005). Retrospective tasks require subjects to recall the initial problem (e.g., pixel density) and their solution (e.g., dense or sparse), and to rate their level of confidence for the answer they provided. Subjects are required to use the confidence-rating scale to retrospectively assess their accuracy. However, like prospective tasks, retrospective tasks introduce an additional layer of complexity in uncertainty monitoring procedures.

Smith and Schull (1989) extended comparative research in uncertainty monitoring tasks by using a psychophysical methodology to test rats in a simultaneous judgment-and-test task. Similar to the pitch discrimination used by Smith, Schull, Strote, McGee, Egnor, and Erb (1995), Smith and Schull used a pitch discrimination task in which rats learned to respond to similarities and differences in tone pairs by pressing levers. Rats also had the option to select an opt-out response by pulling a chain. Over the course of each experimental session, differences in tone pairs were reduced, causing distinctions between identical tone pairs and different tone pairs to be more difficult to determine. As the tones in each pair become more objectively similar, subjects reach a point at which they are equally likely to press the lever associated with the "same" tones in a pair and the lever associated with "different" tones in a pair. The task of categorizing the tones in each pair becomes so difficult that the subjects are reinforced at chance accuracy. If rats' behavior were consistent with that of humans and monkeys in uncertainty monitoring tasks, they would be expected to opt out more frequently on these trials in which they are objectively uncertain (responding at chance) and being reinforced only fifty percent of the time. However, while the rats were able to learn the discrimination task, the results revealed that the rats did not opt out on a large proportion of these objectively uncertain trials. The rats did not appear to use the opt-out response in an adaptive way as humans and monkeys do.

Because these rats did not use the opt-out response often nor did they use it adaptively for objectively uncertain trials, Smith and Schull (1989) questioned whether there was something aversive about using the opt-out. Perhaps the rats were sensitive to the uncertain trials but were simply not motivated to opt out. Therefore, they introduced

a new trial which would, like the uncertain trials previously, result in a fifty percent

reinforcement rate. A high-pitched tone quickly repeating was the new trial type. Unlike

the tone pairs that were the same or different, there was no correct answer for the new

trial type. Regardless of the whether the rats chose "same" or "different", they were

reinforced fifty percent of the time. The number of opt-out responses occurred three

times as often during these chance performance trials that were signaled externally (by a

different stimulus type) as during the uncertain trials. This finding reveals that rats will

use the opt-out option discriminately at a chance reinforcement rate if this rate is signaled

externally.

These experiments suggest that rats do not select opt-out responses in the same

way that humans and rhesus monkeys select them (Smith et al., 1989). Smith and Schull

argue that rats may not experience an internal cue designating uncertainty as humans do

and, thus, they do not use the opt-out response more frequently as the task grows

increasingly difficult. Interestingly, the rats do use the opt-out option for trial types in

which they are reinforced fifty percent of the time and there is no correct answer,

indicating that they are sensitive to chance reinforcement. Specifically, it is the

metacognitive component of the Smith and Schull task that the rats do not perform.

However, Smith and Schull (1989) suggest that a different methodology could

still indicate uncertainty monitoring. Again, it is important to provide nonhuman animals

every opportunity to display uncertainty monitoring in a domain that is most natural for

the animal. What best reveals uncertainty for human participants may not be the same for

nonhuman subjects. Just because rats are not monitoring uncertainty in a given task does

not mean that they are incapable of doing so altogether.

In a more recent attempt at testing the metacognitive capabilities of rats, Foote and Crystal (2007) used an experimental design based on that of Hampton's (2001) research with rhesus monkeys. In this experiment, rats were tested in an auditory duration-discrimination task in which they were first exposed to white noise ranging in length from 2-8 s. Noise durations were logarithmically scaled. Short noises were defined as noises of 2.00, 2.44, 2.97, and 3.62 s in duration, whereas long noises were defined as those of 4.42, 5.38, 6.56, and 8.00 s in duration. The durations closest to the category boundary (3.62 s and 4.42 s) were the most difficult to determine. Initially, correctly categorizing each noise was dependent on trial-and-error learning.

In two-thirds of the trials rats could take the test or opt out. During these trials, left and right nose-poke apertures were illuminated following presentation of noise. One of the nose-pokes was associated with taking the test and resulted in presentation of the retractable levers. In this condition, rats were to then press either a left or right lever to indicate a short or long duration. Correct responses resulted in a food reward of six pellets. Incorrect responses resulted in no reward. The other nose-poke was associated with declining to take the test and did not result in the presentation of levers. Declined trials resulted in an automatic, smaller food reward of three pellets.

In the other one-third of trials, rats were not given the option to opt out. During these trials, only the nose-poke associated with taking the test was illuminated following the stimulus. They were to press either a left or right lever to indicate a short or long duration. Again, correct responses resulted in a food reward of six pellets. Incorrect responses resulted in no reward.

Comparisons in accuracy were made between trials in which the rats had to select an answer (Forced trials) and trials in which the rats could opt out (Choice trials). While only three of eight rats were included in Foote and Crystal's (2007) results, these three declined most frequently on tests in which the discrimination between short and long tones was the most difficult. Accuracy was worse for Forced trials. Foote and Crystal argue that this is the first evidence that rats are capable of metacognition. The rats declined most frequently during difficult trials and their accuracy was worse during Forced trials compared with Choice trials. These results are similar to that of humans and monkeys engaged in uncertainty monitoring tasks.

The outcomes of the two metacognitive studies with rats were significantly different. Those obtained during the prospective task (in which the judgment phase preceded the test phase) used by Foote and Crystal (2007) revealed increased trial opt-outs as difficulty increased and worse performance during Forced trials. Smith and Schull (1989) used a design in which the judgment and test phases were simultaneous. There are important distinctions in the effort required by subjects in the simultaneous versus the prospective tasks. The simultaneous task requires only one response from the subject compared to the prospective task which requires two responses, an indication to opt in or opt out and an answer to the question. For these reasons, one might expect the simultaneous task to have been less taxing than the prospective task. Yet, Smith and Schull's rats did not use the opt-out response more frequently as the task grew increasingly difficult. The unexpected nature and inconsistency in these two sets of results makes the state of metacognitive processes in rats difficult to interpret.

While Foote and Crystal (2007) should be commended for their pioneering work in the area of metacognitive research among rats, several major problems exist with the bold assertions drawn from their experiment. First, the high number of dropouts is a major concern. According to Foote and Crystal, five of the eight rats were unable to learn the contingency of the nose-poke apertures. These five rats did not use the opt-out response adaptively to avoid difficult trials. Therefore, the data for these rats was eliminated. However, it is important to understand why these rats were not learning the contingency of the nose-poke apertures.

Secondly, the procedure used by Foote and Crystal (2007) was a "prospective task" in which the rats were first given a question (a particular duration), then given the option to opt in or out, and finally asked to answer whether the duration was short or long. A prospective task may have been an unnecessarily complicated procedure to use for a fundamental experiment in uncertainty monitoring among rats because it requires an assertion of certainty or uncertainty prior to providing the answer about the duration length.

Another limitation of the Foote and Crystal (2007) experiment was that the lower accuracy for Forced trials compared to Choice trials was only observed during the most difficult trials. Rather than finding a gradual separation in the proportion of correct answers between Choice and Forced trials as difficulty increases, a sharp contrast in the proportion of correct answers between Forced and Choice tests was only found on trials with the greatest stimulus difficulty, or those trials near the subjective middle of the short and long durations. Part of the explanation for this finding is the ceiling effect. The rats' performance was at or near one hundred percent for the majority of the easier trials. To

avoid limited data due to ceiling effects on easier trials, smaller intervals of stimulus difficulty should be used. Further, rats opted out at fairly high rates during easier trials in which their accuracy was high. Because of these inconsistencies, it would be both interesting and scientifically fruitful to investigate more trials at smaller incremental values during the difficult levels.

Additionally, Forced trials went significantly below chance performance at the most difficult level for at least one and possibly two of the three rats included in Foote and Crystal's (2007) results. One would expect the worst performance to be fifty percent, an equal likelihood of choosing the correct or incorrect answer, during the most difficult trials. Instead, one of the rats consistently selected the lever associated with the long tone duration when it was in fact the short tone duration and vice versa. If the metacognitive task was initially learned and performance was reliable, performance should not drop below chance. It is unclear what would cause such a performance pattern. One should note that this behavior cannot be accounted for using a side bias explanation. Bias for either the short or the long lever would result in consistently better performance for either the short or the long tone rather than decreased accuracy for both types of tones. Given that performance was odd on these most difficult trials, one wonders what it means that this was the only trial type for which different accuracies were obtained for Choice versus Forced trials. Clearly, this study has problems that interfere with the claim that this is the first evidence of metacognition in rats that was offered by the authors.

In the current study, the aim was to further explore uncertainty responses in rats. A combination of the strengths between Smith and Schull's (1989) experiment and Foote and Crystal's (2007) experiment was used. In the current study, a tone-length

16

discrimination task similar to that of Foote and Crystal (2007) was used to test

uncertainty in Sprague-Dawley rats. However, the current study did not use a

prospective task in which the rats were required to make a judgment and then take the

test. Like the procedure used by Smith and Schull (1989), the current study prompted the

rats to select an answer following the stimulus. In this case, after rats were exposed to a

stimulus, they were prompted to press one of two levers indicating "short" or "long"

depending on the duration. They also had the option to decline a trial by using a nose-

poke aperture on some trials. The options to opt out of the trial (i.e., break the

photobeam of light in the nose-poke aperture) or to proceed with the trial (i.e., press the

right or left lever) were available simultaneously on these trials. Further, smaller

intervals of stimulus difficulty were presented to avoid limited data due to ceiling effects

on easier trials. In the current study, rats were expected to have the lowest accuracy on

the most difficult test trials in which they could not "opt out." Further, it was expected

that when given the option, rats would "opt out" most frequently on difficult test trials.

This experiment espoused a straightforward design to allow an opportunity for the

rats to display uncertainty monitoring without an intermediate judgment step before the

test phase, which is required during a prospective task. Further, smaller stimulus

increments allowed for a broader variation in accuracy and aided in determining the

causes for particular behavioral patterns. Finally, comparisons in accuracy were also

made for Forced compared to Choice trials.

Methods

Subjects

Two male Sprague-Dawley rats, R01 and R02, served as subjects. Both subjects were approximately 90 days old at the beginning of the experiment and weighed approximately 478 grams. The rats were maintained at 85% of their *ad-libitum* weights through supplementary post-session feeding. *Ad-libitum* water was available in their home cage and testing chambers at all times. The rats were housed together in a polycarbonate cage (480 mm x 270 mm x 220 mm). All procedures were approved by the University of Montana's institutional animal care and use committee (IACUC).

Apparatus

Two identical, sound-attenuated cabinets with a Standard Modular Test Chamber (Med Associates, Inc., ENV-008) were used. Each chamber had two retractable lever presses (Lafayette retractable lever, model number ENV-112CM) positioned along the same plane on one wall. A lever-press was recorded after a 3 mm depression of the lever with a force of 0.20 N. A pellet dispenser was positioned outside the chamber (280 mm above the base) and attached to the food trough (20 mm above the chamber floor and 30 mm from the center of the chamber). Also, a nose-poke aperture was centered 80 mm above the chamber floor and extended 12 mm from the left and right center of the chamber. The nose-poke aperture contained a photobeam that had to be broken in order to detect the presence of the rat's nose. A discriminatory light (110 mcd, orange LED) was added inside the nose-poke to signal activation. A speaker was attached to the test chambers on the wall opposing the lever presses, nose-poke aperture, and pellet

18

dispenser, and a computer delivered auditory stimuli. A computer also recorded lever-press and nose-poke responses.

Procedure

Training procedures were similar to those used by Foote and Crystal (2007). Training and testing procedures were performed seven days a week during the current study. In magazine training, subjects received a sucrose pellet (Dustless Precision pellets, 45 mg; Bio-Serv, Frenchtown, NJ) on a 30 s variable interval schedule for 30 min. Magazine training continued for five days.

Next, lever press pretraining began during which two retractable levers were inserted. The subjects were trained to lever press for 30 min per day. A single sucrose pellet was delivered following orientation toward or movement toward either lever. After five days lever pressing was a reliably established behavior in both subjects. Single lever press training began during which one lever was inserted and a sucrose pellet was delivered based on a single lever press. Then, the previous lever was retracted and either the same lever or the other lever was immediately extended. Again, a single pellet was delivered based on a single lever press. The levers were pseudo-randomly ordered such that no more than three of the same lever was extended in a row. Single lever press training continued until 60 min passed. After two days of single lever press training, the subjects were required to press either lever twice (ratio requirement of two) for the delivery of a sucrose pellet. Single lever press pretraining at a ratio requirement of two continued for approximately three weeks. Both subjects were lever pressing an average of 47 times each training session at the end of this training period.

Next, nose-poke pretraining began. Subjects were trained to break a photobeam of light in a nose-poke aperture centered above the right and left lever presses. Once nose-poking behavior was shaped, nose-poke training began during which subjects were required to break the photobeam to receive a sucrose pellet. Once the photobeam was broken, one sucrose pellet was delivered. This continued until 60 min passed. Nose-poke training continued for five days until the rats were reliably breaking the photobeam with more than seventy nose-pokes per training session.

Duration-discrimination training then began. The subjects were trained to discriminate short and long tone durations. Tones were used in the current experiment in place of white noise used by Foote and Crystal (2007) with the assumption that it might be a more salient stimulus. First, an 1800 Hz pure tone at 70 dB of either 2.0 or 8.0 s was immediately presented. Then, both the right and left levers were inserted. Left and right lever assignment was counterbalanced between subjects. Either a left or right lever press (whichever was associated with the correct response for the short or long tone duration for that particular rat) resulted in a food reward (1 sucrose pellet). Incorrect lever presses did not result in a reward or delay during training. After either a correct or incorrect response, the levers were retracted and an intertrial interval (hereafter ITI) of 45 s occurred. Next, another tone (either 2.0 s or 8.0 s) was presented and the rat was reinforced for pressing the correct lever. Short (2.0 s) and long (8.0 s) tones were pseudo-randomly presented during the remainder of the trials, such that an equal number of both short and long tones were presented and no more than three short or long tones were presented in a row. Duration-discrimination training continued for 60 min. This training lasted for three weeks until the subjects were reliably identifying the durations as

short or long at approximately 80% accuracy. This ended training and a pilot study

began to determine the effectiveness of the training methods.

Pilot Study

During the testing phase of the pilot study, difficult trials were introduced. Trial

difficulty was determined using a similar logarithmic scaling principle used by Church

and Deluty (1977) and Foote and Crystal (2007). However, the task was made more

difficult than Foote and Crystal's in an attempt to reduce or eliminate the ceiling effects

that they found. Eight levels of stimulus difficulty were used between 2.0-8.0 s. Four of

these were short durations and four of these were long durations. The four levels for

short stimuli were (2.00, 2.63, 3.17 and 3.82 s). The four levels for long stimuli were

(4.19, 5.04, 6.06, and 8.00 s). Tone durations closest to the midpoint (i.e., 3.82 and 4.19

s) were most difficult to discriminate (Church & Deluty, 1977; Stubbs, 1976).

Consistent with Foote and Crystal (2007), two-thirds of the trials (16 trials)

included the opt-out option. At the beginning of these trials (Choice trials), an 1800 Hz

pure tone at 70 dB was played. The conclusion of each tone was immediately followed

by the insertion of both right and left levers and the nose-poke aperture was

simultaneously illuminated. The tone duration was pseudo-randomly selected from the

eight different 2.0-8.0 s durations previously described, such that an equal number of

each duration was presented and no more than three short or long tones were presented in

a row. Subjects' responses were measured, including presses on the right or left lever

and photobeam breaks inside the nose-poke aperture. Presses of the "short" lever

contingent with short tones (2.00-3.82 s) and of the "long" lever with long tones (4.19-

8.00 s) were considered correct responses and resulted in a food reward (3 sucrose

pellets). Incorrect responses, lever presses that did not match the tone length, resulted in no reward and a delay (15 s). A third "opt-out" option (the nose-poke) was available. Breaking the photobeam of light in the nose-poke aperture resulted in a lesser food reward (1 sucrose pellet).

Subjects in the current study were rewarded for correct responses with fewer pellets than were the rats in Foote and Crystal (2007). The amount of reinforcement in the current study (3 sucrose pellets for a correct response and 1 sucrose pellet for opting out) was selected due to a concern that the subjects might have opted out too much (Mazur, 1988). One of the subjects in the Foote and Crystal study opted out on more than 40% of trials at the least difficult levels. Such a result makes suspect the claim that this rat was opting out in response to perceived difficulty. In the current task, risk-aversion was indicated by use of the opt-out response (i.e., averting the risk of no reward and a penalty time-out due to an incorrect response). However, an equal but contrasting concern was that the subjects might not have opted out enough for the data to be interpretable. As mentioned in the introduction, this is what happened in the Smith and Schull (1989) study. The use of a nose-poke aperture as the opt-out response was meant to reduce the likelihood that the subjects would not use the opt-out response. Set at an appropriate height, the nose-poke aperture mimicked a natural behavior of rats and was thought to be more likely to encourage responding than the pull chain used by Smith and Schull.

The other one-third of the trials (8 trials) were Forced trials. During the Forced trials the nose-poke aperture was not illuminated and did not function. Subjects were required to select the right or left lever to indicate short or long tone lengths. Sixteen

Choice trials and eight Forced trials were run two times each day for each subject such that forty-eight trials were run for each rat each day. An ITI of 45 s immediately followed the response on each Choice and Forced trial. Accuracy for long and short lever responses was measured for both the Choice and Forced trials. Latency was measured from the onset of the tone until the lever or nose-poke response. Each subject was to have participated in testing until at least 80% consistency in responding across trials was reached.

After three days of testing and no opt-out responses, a nose-poke discrimination program was implemented in which the rats were required to break the photobeam of light inside the nose-poke. The nose-poke was activated on a variable interval schedule of 30 s for twenty trials per day. The nose-poke retraining program was run each day immediately preceding the full testing program. This discrimination program was implemented in an effort to retrain the subjects in the use of the nose-poke following the initial exposure to the nose-poke during training several weeks earlier.

Following two weeks of additional daily testing preceded by the nose-poke discrimination program, no opt-out responses were recorded by either subject. Both rats also exhibited a response bias for the lever associated with the short tones during this two-week testing period. R01 demonstrated 88.8% accuracy for short tones and 63.9% accuracy for long tones during Choice trials; he demonstrated 83% accuracy for short tones and 75% accuracy for long tones during Forced trials. R02 demonstrated 75.8% accuracy for short tones and 72.1% accuracy for long tones during Choice trials; he demonstrated 83.9% accuracy for short tones and 73.2% accuracy for long tones during Forced trials. Several additional program changes were implemented to encourage use of

the nose-poke and to address the response bias. First, one of the eight previously described tones was played prior to the illumination of the nose-poke during the nose-poke discrimination program. Immediately following the tone the nose-poke was illuminated and both levers were also simultaneously extended but not active. Levers remained extended during this training period if pressed, but they did not result in any reward or penalty. Tones were pseudo-randomly presented such that no more than three short or long tones were presented in a row. The addition of the tones during this training program was made due to a concern that previously the only exposure that the subjects had to the tones with the opt-out option available was during testing. This nose-poke discrimination program now exposed rats to each of the eight tones during nose-poke training. Importantly, this training program was not designed to exclusively reinforce a middle tone, nose-poke contingency, as each of the tones was presented during this training.

The second change that was implemented during this time was that Forced and Choice trials were pseudo-randomly presented during testing such that no more than three Forced or Choice trials were presented in a row. Previously, all Forced and Choice trials were presented in blocks with Forced trials preceding Choice trials. Finally, an aversive "buzzer" (70 dB, on/off pattern) was added as a penalty following an incorrect response during the test trials. The penalty buzzer played for 15 s following an incorrect response. The buzzer was followed by a 45 s ITI before the next trial began. The penalty buzzer was introduced to increase learning between correct and incorrect responses.

After two weeks, R01 exhibited only one opt-out response and R02 demonstrated few opt-out responses (7.6% of trials were opted-out by R02). Both rats still exhibited a

response bias for the lever associated with the short tones. R01 demonstrated 85.7% accuracy for short tones and 73.2% accuracy for long tones on Choice trials; he demonstrated 88.7% accuracy for short tones, and 79.5% accuracy for long tones on Forced trials. R02 demonstrated 88.5% accuracy for short tones, and 63.8% accuracy for long tones on Choice trials; he demonstrated 75.9% accuracy for short tones, and 70.7% accuracy for long tones on Forced trials. Thus, two additional measures were taken to encourage opt-out responding and to address the response bias. First, sucrose pellet rewards were reduced to one for both opt-out responses as well as correct responses. This change was implemented as it was determined that the maximum number of reinforcers for each subject was only 56 sucrose pellets per day per subject if he opted-out. However, the subjects' maximum number of reinforcers was 72 sucrose pellets per day if each subject did not opt-out and performed at chance. By changing the sucrose pellet count to a 1:1 ratio for the opt-out and correct responses, this allowed for a maximum of 40 sucrose pellets per day if each subject opted out and 24 sucrose pellets per day if each subject did not opt-out and performed at chance.

The second change that was implemented at this time was a delay allowing for an 8 s total trial period. Thus, a 2 s tone was followed by a 6 s delay and an 8 s tone was followed by no delay. This change was implemented due to a concern that the rats could have been responding to the release of the levers rather than to the tone durations. Behavior changes following a consistent trial duration regardless of tone duration would help to determine whether the subjects were responding based on the length of the tones or the length of the delay period following each tone. Equalizing the total trial period

was an effort to eradicate the inherent differences among trial period based on tone length.

This change created substantial behavior changes in both rats following one week of testing. Both subjects exhibited a strong bias for the lever associated with the long tone durations and each subject's accuracy for the short tone durations decreased considerably. R01 demonstrated 36.5% accuracy for short tones, and 83.8% accuracy for long tones on Choice trials; he demonstrated 39.5% accuracy for short tones, and 69.8% accuracy for long tones on Forced trials. R02 demonstrated 40.5% accuracy for short tones, and 68.3% accuracy for long tones on Choice trials; he demonstrated 27% accuracy for short tones, and 76.8% accuracy for long tones on Forced trials. R02's opt-out rate also increased (32.14% of trials were opted-out by R02) and he frequently attempted to opt-out on Forced as well as Choice trials. R02 exhibited long latencies prior to responding.

One week later, two additional changes were implemented. The ITI was changed from 45 s to 10 s. This was done in an effort to decrease delays and increase motivation. The penalty buzzer was also changed from 15 s to 20 s at this time so that the penalty buzzer would be long in comparison to the ITI. The pilot testing continued for 2.5 weeks. Accuracy remained low following these changes, particularly for short tones. R01 demonstrated 39.5% accuracy for short tones and 75.8% accuracy for long tones on Choice trials; he demonstrated 38.7% accuracy for short tones and 70.6% accuracy for long tones on Forced trials. R02 demonstrated 14.7% accuracy for short tones and 67.9% accuracy for long tones on Choice trials; he demonstrated 26.6% accuracy for short tones and 71.0% accuracy for long tones on Forced trials.

Because of the low number of correct trials for both subjects, it was reasoned that the subjects had not learned the initial tone discriminations. The pilot testing was ceased at this point and an effort was made to retrain the subjects on the tone discriminations to increase accuracy to at least 90%. A series of retraining programs were implemented immediately following the initial testing. First, a tone discrimination program was begun during which 2 and 8 s tones were played in a pseudorandom order in which no more than three of the same tone durations were played in a row. A correct lever press response in which the tone duration matched the corresponding "short" or "long" lever was followed by a single sucrose pellet reward and then a 10 s ITI. An incorrect lever press response was followed by a 20 s penalty buzzer and then a 10 s ITI. This program ran for 60 min for nearly four weeks. R01 responded at 62.3% accuracy for short tones and 83.5% accuracy for long tones, while R02 responded at 81.3% accuracy for short tones and 66.9% accuracy for long tones during the final week of this training phase.

Due to the low accuracies recorded during the unblocked trials, a blocked tone discrimination program was implemented with 2 and 8 s tones played in blocks of ten. For example, a 2 s tone would be presented followed by the presentation of both levers. Following a response, the same tone was presented for nine additional trials. Then the 8 s tone was presented for ten trials. This continued for six rounds (three 2 s rounds and three 8 s rounds) each day. The blocked tone discrimination program ran for four weeks. R01 was reliably responding at 93.8% accuracy for short tones and 89.2% accuracy for long tones, while R02 responded at 93.1% accuracy for short tones and 82.2% accuracy for long tones during the last week of the training period.

Finally, the unblocked tone discrimination program with 2 and 8 s tones was reintroduced for 60 min per day for two weeks. The penalty buzzer was also removed during this time, as there was no evidence that it influenced learning between correct and incorrect responses in either subject. R01 responded at 85.2% accuracy for short tones and 82.3% accuracy for long tones, while R02 responded at 81.3% accuracy for short tones and 73.7% accuracy for the long tones during the final week of this training period. Because performance improved and was consistent for seven days during this training phase, the decision was made to implement the full testing program.

Testing

At this point, the full testing program was introduced during which all tone durations were played and the opt-out option was made available during Choice trials. This program was similar to that previously described in the pilot study with a few important adjustments. A delay allowing for an 8 s total trial period was implemented for each tone duration. All response options, including levers for the Forced trials and levers and the nose-poke for Choice trials, were activated following this 8 s trial period. Correct responses were followed by one sucrose pellet. Incorrect responses were followed by a 20 s timeout instead of a 15 s penalty buzzer. This testing program did not include a penalty buzzer following incorrect responses. Opt-out responses were followed by one sucrose pellet. A 10 s ITI was implemented between trials. Sixteen Choice and eight Forced trials were pseudo-randomly presented such that no more than three of each tone (short or long) and no more than three of each trial type (Choice or Forced) were played in a row. This pattern of pseudo randomly presented tones in two rounds each day. Data collected for each trial included trial type, tone duration, tone type, tone difficulty,

response, response latency, and outcome. This testing phase continued for seven weeks until stability in the number of opt-out responses was reached for each subject.

Results

First, it was expected that the rate of opt-out responses would increase as the difficulty of the stimulus discrimination increased during Choice trials. Difficulty levels were collapsed across tone durations for all analyses with Level 1 indicating the least difficult tone durations (2.00 and 8.00 s) and Level 4 representing the most difficult tone durations (3.82 and 4.19 s). An alpha level of .05 was used for all statistical tests. No analysis was performed for opt-out responses for R01, as he never selected the opt-out option during testing. A chi-square test of independence was performed to examine the relation between opt-out responses and difficulty level for R02. The relation between these variables was not significant, X^2 (3, $N = 1$) = 6.81, $p > .07$. Opt-out responses did increase as difficulty level increased for R02, but not significantly (see Figure 1).

Figure 1.

Opt-out Responses by Level of Difficulty for R02
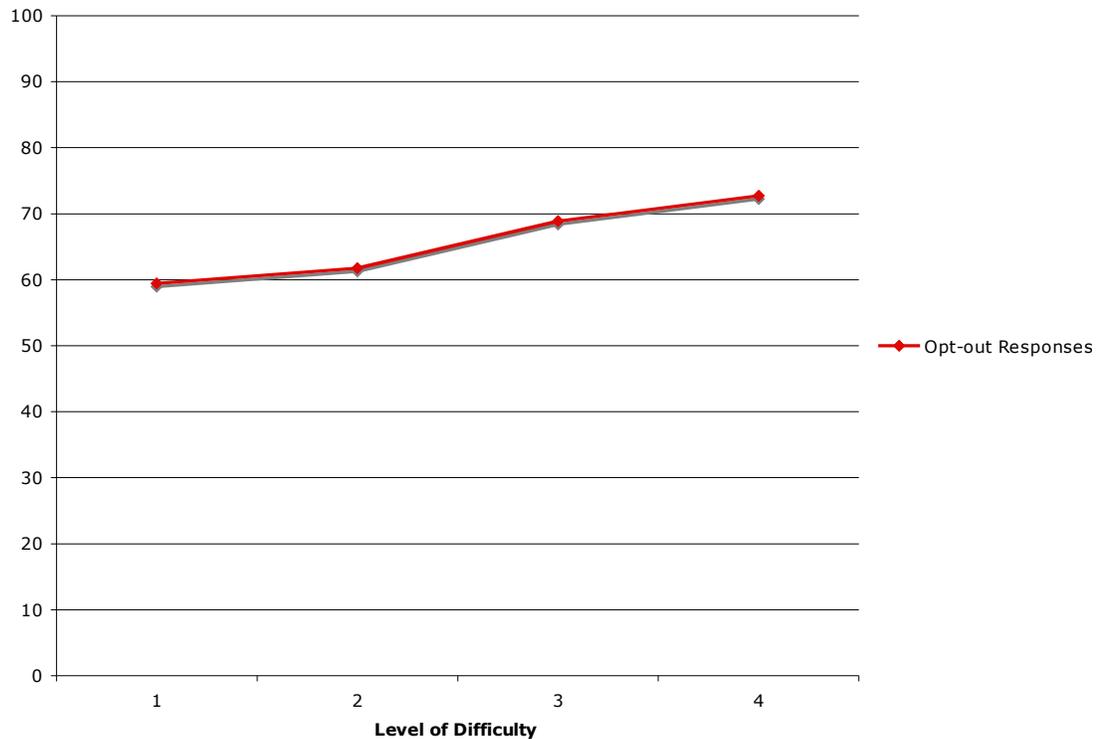
Secondly, accuracy was expected to decline as difficulty of the stimulus

discrimination increased. A chi-square test of independence was performed to examine

the relation between accuracy and difficulty level for each rat within choice and forced

trials. Choice trial accuracy significantly decreased as the level of difficulty increased for

both subjects (R01 Forced trials, X^2 (3, $N = 1$) = 117.55, $p < .00$; R01 Choice Trials, X^2 (3,

$N = 1$) = 37.68, $p < .00$ (see Figure 2); R02 Choice trials, X^2 (3, $N = 1$) = 28.65, $p < .00$;

R02 Forced trials, X^2 (3, $N = 1$) = 8.77, $p < .00$ (see Figure 3).

Figure 2.
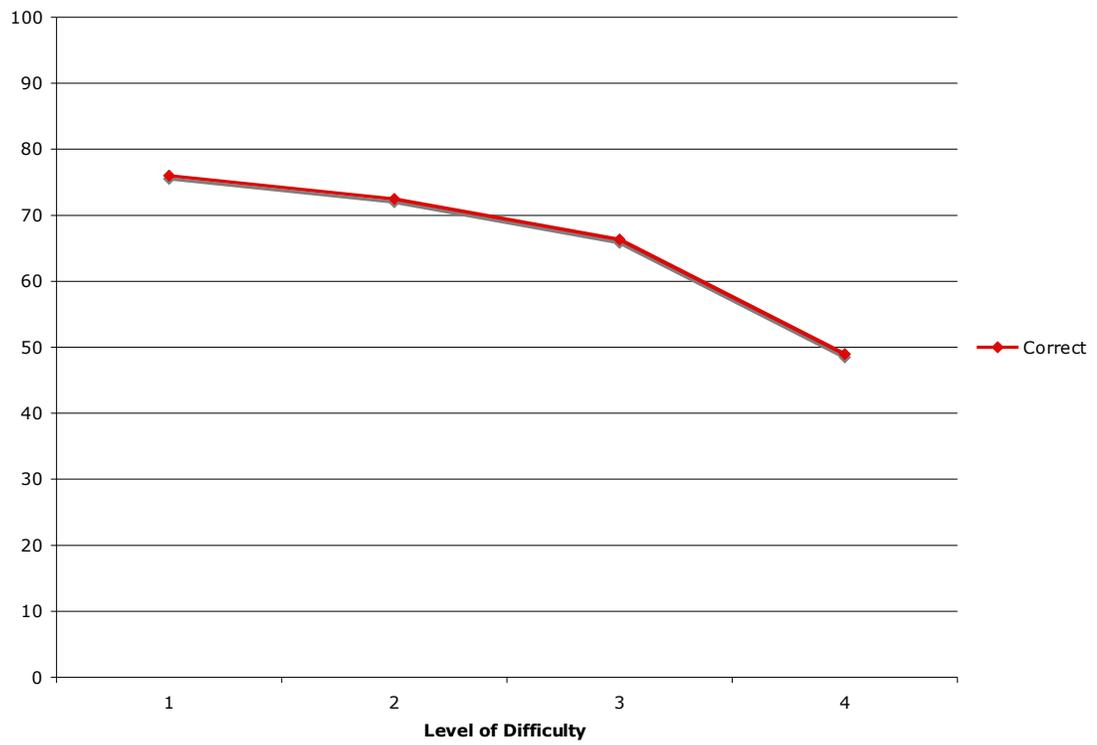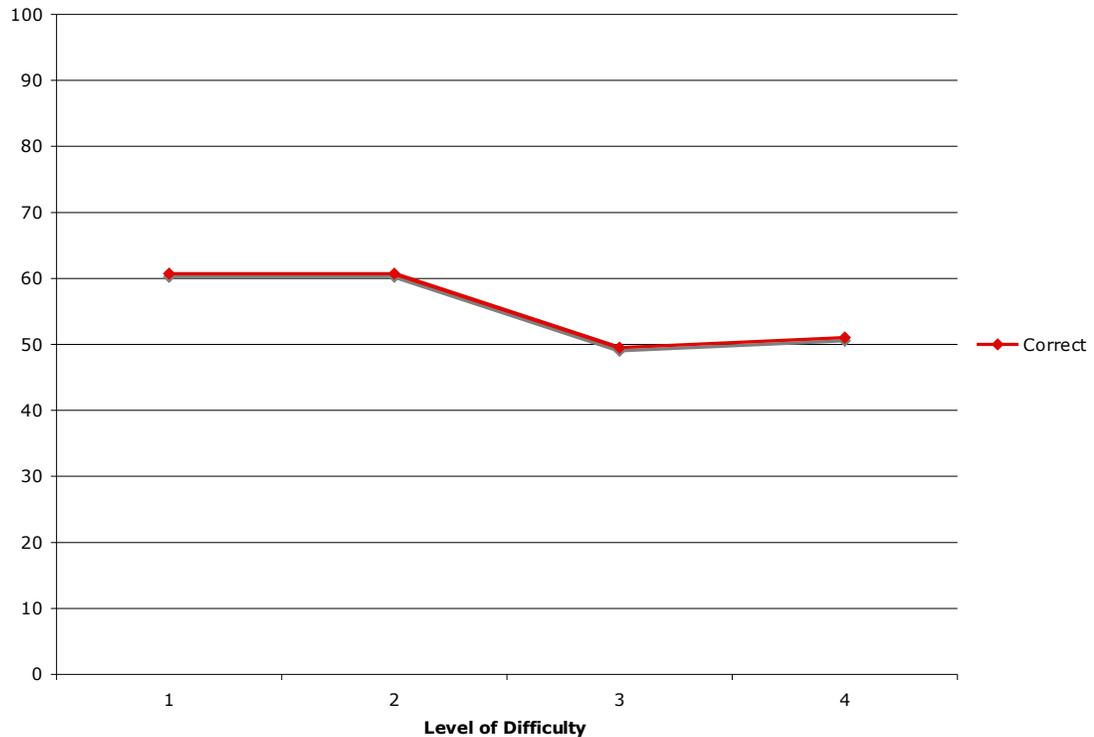Percentage of Correct Responses by Level of Difficulty, R01 Forced Trials

Figure 3.
Percentage of Correct Responses by Level of Difficulty, R02 Forced Trials

Finally, accuracy was expected to be higher on Choice trials compared to Forced trials. Subjects should have opted-out on the trials in which they were more likely to err, increasing the accuracy during Choice trials because it did not include the opted-out trials. A one-tailed z-test for two proportions was performed to determine whether the accuracy of the Choice and Forced groups were significantly different at each of the four levels of difficulty. A significant difference was found for the accuracy of Choice and Forced trials for R01 only for difficulty Level 1, $z = 1.91$, $p < .05$. No significant difference was found for the accuracy of Choice and Forced trials for Level 2, $z = 0.10$, $p > .05$; for Level 3, $z = 0.62$, $p > .05$; or for Level 4, $z = 0.14$, $p > .05$ (see Figure 4). One would not expect a difference in accuracy between Choice and Forced trials since R01

did not opt-out. Accuracy was determined by removing opt-out responses from the total number of Choice trials for R02. A significant difference was found for the accuracy of Choice and Forced trials for R02 for each level of difficulty, with Level 1, $z = 5.12$, $p < .05$; Level 2; $z = 3.26$, $p < .05$; Level 3, $p < .05$, $z = 2.67$; and Level 4, $z = 1.70$, $p < .05$ (see Figure 5). R01's accuracy was only statistically significant between Choice and Forced trials during the least difficult levels. R02, however, displayed significantly better performance on Choice compared to Forced trials on all levels of difficulty.

Figure 4.
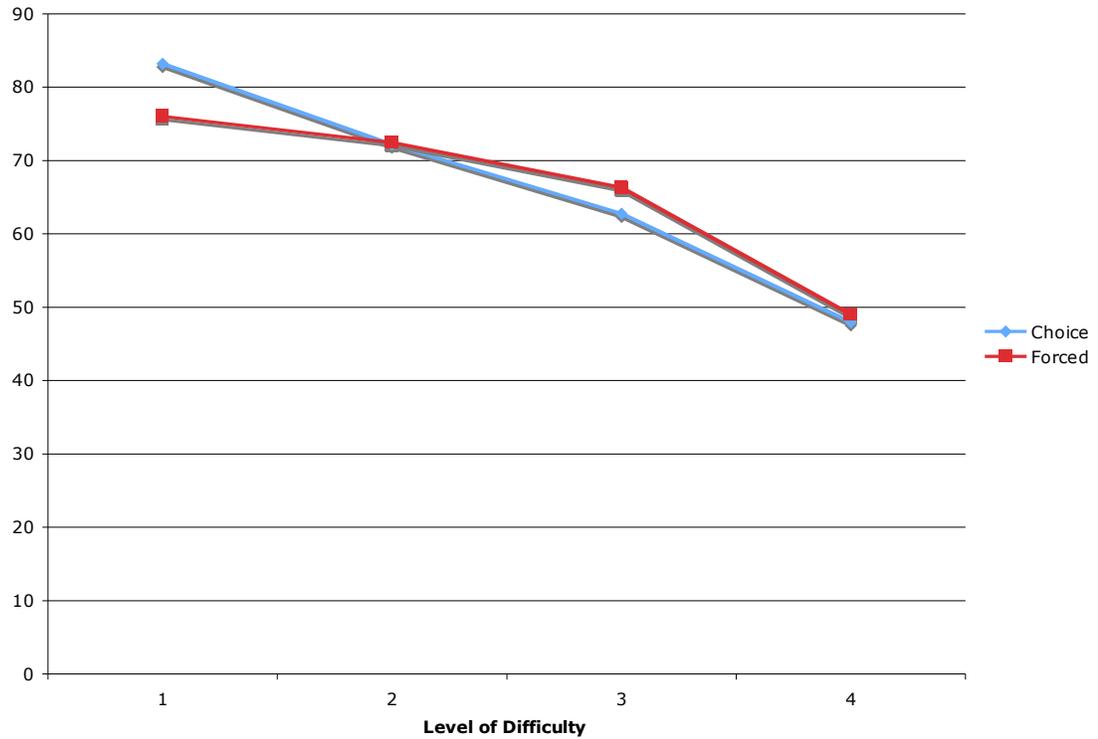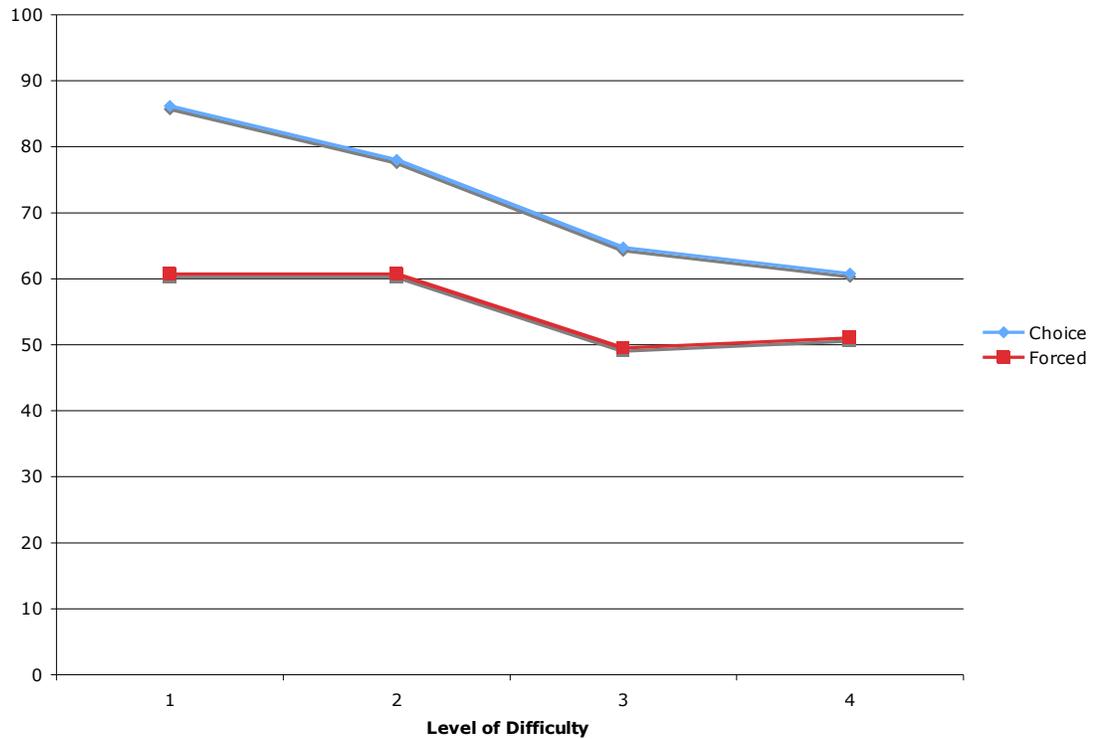Percentage Correct for Choice and Forced Trials for R01

Figure 5.
Percentage Correct for Choice and Forced Trials for R02

Discussion

This experiment provides at least some evidence that one rat knows when it does not know the correct response on a duration-discrimination task. If the rats knew when they did not know, they should have opted out on the trials in which they were more likely to err. Also, the accuracy of Choice trials should be higher than the accuracy of the Forced trials if the rats knew when they did not know. Although the frequency of R02's opt-out responses did not increase significantly from one level of difficulty to the next, overall, he did show a trend toward increased opt-out responses as difficulty increased. Also, R02's accuracy for Choice trials was higher than his accuracy for Forced trials at each level of difficulty. Forced trial accuracies included trials with tone durations that were difficult and likely would have been declined if the opt-out option had been

34

available as it was during Choice trials. This provides evidence for uncertainty

monitoring in one rat.

R01 did not show evidence of uncertainty monitoring, as he did not opt-out during

testing. This could suggest that he did not know when he did not know. Alternatively,

this behavior might suggest that the trials were not sufficiently difficult enough to

encourage R01 to opt-out. This argument would logically follow if R01 had been

capable of responding accurately on most trials, regardless of increasing difficulty levels.

In fact, R01's accuracy was higher than R02's accuracy for all but the most difficult trial

level during Forced trials. Yet, R01's accuracy decreased as trial difficulty increased for

both Forced and Choice trials. His accuracy at the most difficult level was at chance

performance for both Forced and Choice trials. Therefore, regardless of his improved

performance relative to R02, it would have been adaptive for him to use the opt-out

response as trial difficulty increased.

One explanation for R01's lack of opt-out responses during testing could be that

he did not learn the nose-poke behavior during training. However, he quickly learned the

contingency for the initial nose-poke training, as well as the nose-poke discrimination

program used prior to testing during the pilot study. R01 simply did not apply the nose-

poke response learned during training to the testing trials. One possibility as to why R01

did not opt-out during testing is that he responded only to the levers when they were

made available. The likelihood of this explanation is increased by the fact that the levers

produce noise when extended while the nose-poke is inaudible when activated. Several

measures were taken during pilot testing to reduce this inconsistency. First, a

discriminatory light (110 mcd, orange LED) was added above the nose-poke in each

chamber in addition to the same light that was originally added to the inside of the nose-poke. Both lights were illuminated upon activation of the nose-poke. Secondly, the nose-poke discrimination program introduced immediately before the pilot tests was modified to include extended and inactive levers. Once a tone was played, both levers were extended and the nose-poke was simultaneously activated. No reward or penalty was delivered until the photobeam inside the nose-poke was broken. These efforts had no impact on R01's tendency to lever-press when presented with the opportunity. While it is unclear what motivated this behavior, future research efforts would likely benefit from reducing the difference in saliency between the onset of the opt-out and opt-in options. For example, two identical auditory cues or visual cues should signal the onset of both options.

This experiment was designed to further explore comparative research in uncertainty monitoring in rats. It was hoped that the results would clarify and extend the findings of previous metacognition experiments in rats by Smith and Schull (1989) and Foote and Crystal (2007). One method that was planned to meet this goal was to increase the number of subjects to avoid limited results from a small sample size. However, due to the difficulty in reliably establishing nose-poke responses and tone discrimination during testing, only two subjects were included in this experiment. There is no evidence to suggest that these two subjects were exhibiting unique behavior. In fact, Foote and Crystal (2007) faced similar challenges in that five of their eight rats "rarely declined to take the duration test" likely because they "failed to learn the experimental contingency of the nose-poke aperture."

Another strategy implemented to clarify and extend previous findings was the use of a simultaneous judgment-and-test design. The simultaneous task is a more straightforward approach requiring only one step in which both the judgment and test are presented together, whereas the prospective task requires an initial judgment about future behavior prior to giving an answer. The simultaneous task is thought to be a less difficult task, although as previously indicated, the subjects in this experiment had numerous problems learning the duration-discrimination task and effectively utilizing the opt-out response. Several efforts were taken to improve accuracy on the duration-discrimination task following the pilot study. A series of tone discrimination programs were initiated, including unblocked 2.00 and 8.00 s tone durations, blocks of tones of the same duration, and again, unblocked tones of the same duration. This additional training occurred daily over a nine-week period, improving performance to an average of 85-90% accuracy.

Another method in this experiment was to use more difficult stimuli to avoid the ceiling effects found by the Foote and Crystal experiment. Intermediate tones were introduced during testing to increase the level of difficulty during the task. It was hypothesized that accuracy would increase as the level of difficulty of the task increased. R01 and R02 both demonstrated decreased performance as the level of difficulty increased for both Forced and Choice trials. Importantly, while both subjects' accuracy decreased on more difficult trials, the performance trends varied between the two subjects. R01's performance graph shows a linear decline by difficulty level, while R02's performance graph only declines between difficulty Levels 2 and 3. The performance of both of the subjects in this study differs from Foote and Crystal's (2007)

results, such that performance declined throughout and during increasing levels of difficulty rather than only declining during the most difficult level.

An additional limitation with this experiment could have been that the intermediate tones were not introduced during pretraining. This method was used to enhance tone discrimination between the short and long anchor tones. However, the first time the rats were exposed to the intermediate tones was during the initial testing phase and this was also the first time that they had been exposed to a tone followed by the availability of the opt-out option. This could have influenced the way that the intermediate tones were categorized and the behavioral responses to them. This training may have created a contingency in which the opt-out option became the middle response rather than an uncertainty response (e.g., Beran, Smith, Coutinho, Couchman, & Boomer, 2009). Further studies may benefit from a pretraining period during which the rats are exposed to intermediate tones. It is important to note, however, that opt-out responses particular to intermediate categories should be avoided during pretraining so that the animal is not simply reenacting previously reinforced behavior (Terrace & Son, 2009).

A final, and most significant, improvement upon the Smith and Schull (1989) experiment was to compare accuracy on trials in which the rats could opt out and those in which they could not. Unlike the Smith and Schull (1989) experiment, comparisons in accuracy between Choice and Forced trials during the current study allowed for a better understanding of how the opt-out option was being used. If rats were using the opt-out option adaptively and were opting out on trials in which they did not know the answer, accuracy was expected to be higher for the Choice compared to the Forced trials. If the rats were not using the opt-out option in an adaptive way, no significant differences

would have been expected between the Choice and Forced trials because opted-in and opted-out trials would be selected at random. A significant difference was found between Choice and Forced trials at each level of difficulty for R02. R01 did not opt-out during testing although he did display a difference between Choice and Forced trials only at the least difficult trial level. This finding is not consistent with R01's other responses during testing and is unlikely to hold theoretical significance.

One challenge that existed in this study was the use of a duration discrimination task. Duration tasks present particular difficulties during comparative research. One problem is that duration tasks such as the one used in this experiment are temporal, memory-based tasks, requiring subjects to recall the onset and offset of the tone and to make a comparison between the two times. Previous studies have suggested that rats experience episodic-memory similarly to humans (Babb & Crystal, 2005a, 2005b; Babb & Crystal, 2006; Crystal, 2006). More recent evidence suggests that while rats can successfully track short and long time intervals, they may differ significantly from humans in how they solve such tasks. For example, there is evidence to suggest that rats rely on cues regarding how long ago an event occurred rather than mentally time traveling to determine when an event occurred as humans and perhaps some other animals, such as scrub jays, do (e.g., Clayton, Yu, & Dickinson, 2001; Roberts, Feeney, MacPherson, Petter, McMillan, & Musolino, 2008). This suggests that rats may rely on other cues, such as elapsed time, to solve these tasks (Friedman, 1993). Frequency discrimination, which is a basic component of the auditory system of mammals, may provide a simpler perceptual problem-solving task that could be used during uncertainty

monitoring tasks in nonhuman species without the complication introduced by temporal processing tasks (e.g., Talwar & Gerstein, 1998).

An additional challenge during this study was determining the amount of reward to be delivered for opt-in and opt-out responses. The sugar pellet reward was reduced to one pellet for both a correct response and an opt-out response following initial testing. By changing the sucrose pellet count to a 1:1 ratio for the opt-out and correct responses, this allowed for a maximum of 40 sucrose pellets per day if each rat opted out and 24 sucrose pellets per day if each rat did not opt-out and performed at chance. Reducing the correct response reward quantity was a measure taken to encourage opt-out responses as there was previously no net reward to be gained by opting-out. However, one could argue that this methodological change reduces the likelihood of a cognitive explanation for why the rats might use the opt-out response at all. Previous studies involving comparative metacognition have been criticized when subjects are reinforced for opt-out responses or when opt-out responding increases the frequency of reward (Beran, Smith, Coutinho, Couchman, & Boomer, 2009; Terrace & Son, 2009). This methodological approach may create a middle response rather than an uncertain response and lends itself to an associative explanation rather than a cognitive one. Effort should be made in future comparative metacognition research to merge learning research in behavioral economics with uncertainty monitoring tasks. Some success has already occurred on this front in studies that have utilized post-decision wagering to uncover information regarding certainty and uncertainty about a question posed during a cognitive or perceptual task (Kepecs, Uchida, Zariwala, Mainen, 2009; Koch & Preuschoff, 2007; Persuad, McLeod, & Cowey, 2007).

The goal of this study was to provide clarification to previous findings and to

discover further evidence about uncertainty monitoring in rats. This experiment provides

additional evidence that at least one rat may know when it does not know the correct

response to a duration-discrimination task. Such evidence lends greater understanding to

the evolutionary development of metacognition and reveals compelling insights into the

phylogeny of at least one aspect of metacognitive processing. Future research should

consider the challenges in comparative uncertainty monitoring tasks previously

discussed, including the behavioral implications of the amount of reward disseminated,

the confounds associated with the use of a temporal-based task, and the complexity of a

psychological middle.

References

Babb, S. J. & Crystal, J. D. (2005). Discrimination of what, when, and where: Implications for episodic-like memory in rats. *Learning & Motivation*, 36(2), 177-189.

Babb, S. J. & Crystal, J. D. (2006). Discrimination of what, when, and where is not based on time of day. *Learning & Behavior*, 34(2), 124-130.

Babb, S. J. & Crystal, J. D. (2006). Episodic-like memory in the rat. *Current Biology*, 16, 1317-1321.

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(3), 371-381.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Carruthers, P. (2008). Meta-cognition in animals: A skeptical look. *Mind & Language,* 23, 58-89.

Chua, E. F., Schacter, D. L., & Sperling, R. A. (2009). Neural correlates of metamemory: A comparison of feeling-of-knowing and retrospective confidence. *Journal of Cognitive Neuroscience*, 21, 1751-1765.

Church, R. M. & Deluty, M. Z. (1977). Bisection of temporal intervals. *Journal of Experimental Psychology: Animal Behavioral Processes*, 24, 151-162.

Clayton, n. S., Yu, K. S., & Dickinson, A. (2001). Scrub jays (*Aphelocoma coerulescens*) form integrated memories of the multiple features of caching episodes. *Journal of Experimental Psychology: Animal Behavior Processes*, 27, 17-29.

Crystal, J. D. (2006). Time, place, and content. *Comparative Cognition and Behavior Reviews*, 1, 53-76.

Fernandez-Duque, D., Baird, J. A., Posner, M. I. (2000). Executive attention and metacognitive regulation. *Conscious Cognition*, 9, 288-309.

Flavell, J. H. (1976). Metacognitive aspects of problem solving. In L. B. Resnick (Ed.), *The Nature of Intelligence* (pp. 231-236). Hillsdale, NJ: Erlbaum

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, Motivation, and Understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

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

Friedman, W. J. (1993). Memory for the time of past events. *Psychological Bulletin*, 133, 44.

Gallo, D. A., Kensinger, E. A., Schacter, D. L. (2006). Prefrontal activity and diagnostic monitoring of memory retrieval: fMRI of the criterial recollection task. *Journal of Cognitive Neuroscience*, 18, 135-148.

Hampton, R. (2001). Rhesus monkeys know when they remember. *Proceedings of the National Academy of Sciences*, USA, 98, 5359-5362.

Hart, J. T. (1965). Memory and the feeling-of-knowing experience. *Journal of Educational Psychology*, 56, 208-216.

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

James, W. (1890). Principles of Psychology. New York: Henry Holt & Co.

Janowsky, J. S., Shimamura, A. P., & Squire, L. R. (1989). Memory and metamemory: Comparisons between patients with frontal lobe lesions and amnesic patients. *Psychobiology*, 17, 3-11.

Kepecs, A., Uchida, N., Zariwala, H. A., Mainen, Z. F. (2008). Neural correlates, computation and behavioral impact of decision confidence. *Nature*, 455, 227-231.

Koch, C. & Preuschoff, K. (2007). Betting the house on consciousness. *Nature Neuroscience*, 10, 140-141.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review*, 100, 609-639.

Koriat, A. (1994). Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 115-135). Cambridge, MA: MIT Press.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence

    for the accessibility model. *Journal of Experimental Psychology: General*, 124,

    311-333.

Kreutzer, M. A., Leonard, C., & Flavell, J. H. (1975). An interview study of children's

    knowledge about memory. *Monographs of the Society for Research in Child*

    Development, 40, 1.

Markman, E. M. (1979). Realizing that you don't understand: Elementary school

    children's awareness of inconsistencies. *Child Development*, 50, 643-655.

Mazur, J. E. (1988). Choice between small certain and large uncertain reinforcers.

    Animal Learning and Behavior, 16, 199-205.

McGuinness, C. (1990). Talking about thinking: The role of metacognition in teaching

    thinking. *Lines of Thinking*, 2, 301-312.

McGuinness, C. (2005). Teaching thinking: Theory and practice. *Pedagogy – Learning*

    for Teaching: BJEP Monograph Series, 11, 107-126.

Miner, A. C. & Reder, L. M. (1994). A new look at feeling of knowing: Its

    metacognitive role in regulating question answering. In J. Metcalfe & A. P.

    Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 47-70).

    Cambridge, MA: MIT Press.

Nelson, T. O. & Narens, L. (1990). Metamemory: A theoretical framework and new

    findings. *The Psychology of Learning and Motivation, 26, 125-141*.

Nelson, T. O., Narens, L., & Dunlosky, J. (2004). A revised methodology for research

    on metamemory: Pre-judgment recall and monitoring (PRAM). *Psychological*

    Methods, 9, 53-69.

Persuad, N., McLeod, P., & Cowey, A. (2007). Post-decision wagering objectively

    measures awareness. *Nature Neuroscience*, 10, 257-261.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?

    Behavioral and Brain Sciences, 4, 515-526.

Reder, L. M. (1994). Difference research programs on metacognition: Are the

    boundaries imaginary? *Learning and Individual Differences*, 8, 383-390.

Roberts, W. A., Feeney, M. C., MacPherson, K., Petter, M., McMillan, N., & Musolino,

    E. (2008). Episodic-like memory in rats: Is it based on when or how long ago?

    Science, 320, 113-115.

Seth, A. K., Edelman, D. B., & Baars, B. J. (2004). Let's not forget about sensory

    consciousness. *Behavioral and Brain Sciences*, 27, 601-602.

Shields, W. E. (1999). *Nonverbal judgments-of-knowing by humans and Rhesus*

    monkeys. Unpublished doctoral dissertation, State University of New York at

    Buffalo.

Shields, W. E., Smith, J. D., Guttmannova, K., & Washburn, D. A. (2005). Confidence

    judgments by humans and rhesus monkeys. *Journal of General Psychology,* 132,

    165-186.

Shields, W. E., Smith, J. D., & Washburn, D. A. (1997). Uncertain responses by humans

    and rhesus monkeys (*Macaca mulatta*) in a psychophysical same-different task.

    Journal of Experimental Psychology: General, 126, 147-164.

Shimamura, A. P. (1994). The neuropsychology of memory. In J. Metcalfe & A. P.

    Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 253-276).

    Cambridge, MA: MIT Press.

Shimamura, A. P. (2000). Toward a cognitive neuroscience of metacognition. *Consciousness and Cognition*, 9, 313-323.

Smith, J. D. (2009). The study of animal metacognition. *Trends in Cognitive Sciences*, 13(9), 389-396.

Smith, J. D. & Schull, J. (1989). A failure of uncertainty monitoring in the rat. (unpublished data).

Smith, J. D., Schull, J., Strote, J., McGee, K., Egnor, R., & Erb, L. (1995). The uncertain response in the bottlenosed dolphin (*Tursiops truncatus*). *Journal of Experimental Psychology: General,* 124, 391-408.

Smith, J. D., Shields, W. E., Allendoerfer, K. R., & Washburn, D. A. (1998). Memory Monitoring by Animals and Humans. *Journal of Experimental Psychology: General*, 127, 227-250.

Smith, J. D., Shields, W. E., & Washburn, D. A. (2003). The comparative psychology of uncertainty monitoring and metacognition. *Behavioral and Brain Sciences,* 26, 317-373.

Sole, L. M., Shettleworth, S. J., & Bennett, P. J. (2003). Uncertainty in pigeons. *Psychonomic Bulletin & Review,* 10, 738-745.

Stubbs, D. A. (1976). Scaling of stimulus duration by pigeons. *Journal of the Experimental Analysis of Behavior*, 26, 15-25.

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

Terrace, H. S. & Son, L. K. (2009). Comparative metacognition. *Current Opinion in Neurobiology*, 19, 67-74.

Van der Werf, Y. D., Witter, M. P., & Groenewegen, II, J. (2002). The intralaminar and

      midline nuclei of the thalamus: Anatomical and functional evidence for

      participation in the processes of arousal and awareness. *Brain Research: Brian*

      Research Reviews, 39, 107-140.

Washburn, D. A., Smith, J. D., & Shields, W. E. (2006). Rhesus monkeys *(Macaca*

      mulatta) immediately generalize the uncertain response. *Journal of Experimental*

      Psychology: Animal Behavior Processes, 32, 185-189.

Wellman, H. M. (1977). Preschoolers' understanding of memory-relevant variables.

      Child Development, 48, 1720-1723.