Published in final edited form as:
Comp Cogn Behav Rev. 2009 January 1; 4: 17–28.

Multiple demonstrations of metacognition in nonhumans: Converging evidence or multiple mechanisms?

Robert R. Hampton
Emory University

Abstract

Metacognition allows one to monitor and adaptively control cognitive processes. Reports from the last 15 years show that when given the opportunity, nonhuman animals selectively avoid taking difficult tests of memory or perception, collect more information if needed before taking tests, or “gamble” more food reward on correct than on incorrect responses in tests of memory and perception. I review representative examples from this literature, considering the sufficiency of four classes of mechanism to account for the metacognitive performance observed. This analysis suggests that many of the demonstrations of metacognition in nonhumans can be explained in terms of associative learning or other mechanisms that do not require invoking introspection or access to private mental states. Consideration of these accounts may prompt greater appreciation of the diversity of metacognitive phenomena and may inform theoretical positions about the nature of the mental representations underlying metacognition

Keywords

awareness; cognitive control; confidence; consciousness; declarative; explicit; introspection; memory; memory monitoring; metacognition; metamemory; perception; self-awareness; self-control; self-regulation; uncertainty

Metacognition: A broad definition

Metacognition allows one to monitor and adaptively control cognitive processing (e.g. Nelson, 1996; Nelson & Narens, 1990). For example, a student might improve her grade by dedicating more of her study effort to the longest text-book chapters and the most difficult topics on an upcoming exam. She might re-study the definitions of terms she finds she forgot after a single study session. Finally, during the exam she might skip questions for which she is unsure of the answer, returning to them only after first answering questions for which she is confident. In each case, our student has assessed the difficulty faced in learning or performing and has adjusted her behavior appropriately.

Our student’s behavior in each of the above examples deserves the label metacognition, at least as it is broadly defined (Flavell, 1979). Demonstrations of metacognition in the laboratory must meet four criteria:

1. We must specify a primary, objectively observable behavior that can be scored for accuracy or efficiency. Accuracy might be assessed as questions answered correctly, while efficiency could be assessed as time taken to learn all the assigned material.

Correspondence concerning this article should be addressed to: Robert R. Hampton, Department of Psychology, 532 Kilgo Circle, Emory University, Atlanta, GA, USA. robert.hampton@emory.edu.

2. There must be variation in the accuracy or efficiency of the primary behavior. Variation in performance is necessary in order to allow assessment of the correlation between the primary behavior and the secondary behavior (described in 3, below).
3. We must specify a secondary, objectively observable behavior that can be used to infer monitoring or regulation of cognition underlying the primary behavior. Monitoring of knowledge might be indicated by skipping questions for which the subject is unsure of the answer, while regulation might be indicated by subjects adjusting time spent studying to match the difficulty of the material.
4. There must be an explicit assessment of whether the primary and secondary behaviors are correlated. For example, were the questions that the subject skipped indeed ones for which he did not know the answer? Was study time adjusted appropriately to increase efficiency of learning? This correlation can be assessed most powerfully when the subject's knowledge is experimentally manipulated and their knowledge state can therefore be confidently known. For example, do subjects skip questions relating to information that has not been provided during training, while answering questions for which they have been trained on the correct answer?

Given the objective nature of the four criteria outlined above, it is possible to devise tests of metacognition for nonhuman animals. In fact, a substantial literature has developed over the last 15 years demonstrating that several nonhuman species clearly meet all four of these criteria. In perceptual tests, monkeys, dolphins, and rats have been shown to either decline difficult trials or make accurate post-trial confidence judgments (e.g. Foote & Crystal, 2007; Kornell, Son, & Terrace, 2007;J. D. Smith, et al., 1995). Monkeys performing memory tests have also been shown to adaptively decline difficult tests and to make adaptive confidence judgments about previous performance (Hampton, 2001, 2005; Kornell, et al., 2007; J.D. Smith, Shields, & Washburn, 1998). Pigeons have not reliably shown similar results (Inman & Shettleworth, 1999; Sole, Shettleworth, & Bennett, 2003; Sutton & Shettleworth, 2008). Note that meeting the four criteria required for metacognition does not by itself specify what particular mechanism underlies the correlation between the primary performance and the secondary metacognitive response. Metacognitive performance can potentially be achieved through a variety of mechanisms, some of which may be entirely consistent with traditional views of nonhuman cognition and others that might call for re-evaluation of the richness of nonhuman cognition.

Private and public mechanisms for metacognition

Metacognition in humans is often associated with conscious awareness of one's own cognitive states (e.g. Koriat, 1996; Nelson, 1996) and is therefore presumed to reflect private monitoring of those states. I will argue that it is theoretically important to distinguish between private and public mechanisms for metacognition. *Private* mechanisms are those by which cognitive control is contingent on the privileged access the subject has to their own cognitive states. In the case of *public* mechanisms, adaptive cognitive control is based upon the use of publicly available information such as the perceivable difficulty of a problem or the subject's reinforcement history with particular stimuli. Contrast the following two situations requiring a metacognitive judgment: 1) a colleague asks whether you remember the title of B. F. Skinner's first book, 2) a friend asks whether you can answer a question his six year old has about psychology. In the first case, you would surely check the contents of your memory and determine whether you can retrieve a memory of the book title. Your metacognitive judgment would therefore depend on your success or failure at privately retrieving the relevant explicit memory, a cognitive state to which you, as the one doing the remembering, have privileged access. In the second case, your friend has not even asked you to retrieve a specific memory. If you are an expert in Psychology, you might feel confident (probably correctly) that you can answer the question of a six year old. However, your confidence would not depend on a private evaluation of your memory. Instead, your confidence would depend on your history of expertise, your past ability to answer such questions, and your assessment of the intellectual capacity of six year olds – all publicly available information. It is significant that, in the second case, your friend's judgment about your ability to answer correctly would be about as accurate as your own. This would not be true if you were introspectively accessing a specific explicit memory, in which case you as the introspecting individual would have a distinct advantage over others in accurately estimating your knowledge. Similarly, students can be trained to allocate more study effort to complex topics or to courses that will require memorization of many details. Such adaptive behavior can be guided by an assessment of the material to be studied that could be carried out equally well by the learner *or by another person*. The cues that indicate the difficulty of the material (complexity, number of terms to memorize, etc.) are publicly observable. Thus, the observation of adaptive cognitive control should not be uncritically equated with private mechanisms. Adaptive control of study effort does not require introspective access to information that only the learner or performer possesses.

To understand the mechanisms of metacognition in nonhumans we will have to do more than demonstrate adaptive cognitive control. We will have to develop experimental procedures that allow us to specify what information subjects use to assess their ability to learn or perform, and how they use that information. A fear probably shared by all investigators of nonhuman metacognition is that we are misinterpreting "Clever Hans" type phenomena in which apparently impressive cognitive feats can be accomplished by established "simple" mechanisms (e.g. Roberts, 1998, p. 9–11; S.J. Shettleworth, 1998, p. 363). Studies of humans provide additional reasons for caution in ascribing complex mechanisms, particularly introspection, solely on the basis of the complexity of behavior. Humans can accomplish a great deal of learning and performing without conscious awareness or introspection, for example, classical conditioning (Clark & Squire, 1998), skill learning (Cohen, Eichenbaum, Deacedo, & Corkin, 1985; Knowlton, Ramus, & Squire, 1992; Knowlton & Squire, 1993), and priming (Hamann & Squire, 1997; Tulving & Schacter, 1990).

Lumping all types of successful metacognition under a single descriptive term may obscure important differences in metacognitive function (e.g. Hampton, 2003). Because public mechanisms of metacognition depend on publicly observable information, their operation can likely be explained in terms traditional to animal learning and comparative cognition. By contrast, evidence for private mechanisms involving some type of introspection might require that we extend our understanding of what nonhuman animals perceive to include some of their own cognitive states. Perceived cognitive states may enter into associations and control behavior according to the same rules that govern these processes with respect to overt stimuli; there is no *a priori* reason to suspect that different rules apply. That is, whether a discriminative stimulus is an overt light or a private assessment of memory may have little impact on the way it controls behavior. But control of behavior by monitoring of private cognitive states may support flexible and adaptive behavior that would not be possible using publicly available information (Hampton, 2005).

Four classes of stimuli sufficient for metacognitive control

Most or all cases of nonhuman metacognition may be adequately accounted for by public mechanisms. In the following cases, it is not possible to determine with confidence what stimuli indeed control the observed metacognitive responses. Because we cannot obtain from nonhumans the verbal reports that constitute part of the evidence for private introspective metacognition in humans, we can only infer private metacognition in nonhumans by excluding likely public mechanisms. The procedures used in published reports differ in the extent to which they exclude classes of public mechanisms for metacognition. Below, I describe four classes of mechanism for metacognition. This list is unlikely to be exhaustive; I hope that it is representative.

Environmental cue associations

Some stimuli are more difficult to discriminate or remember than are others and some test conditions are more challenging than are others. Stimuli that are close together on a continuum are more difficult to discriminate than are those that are far apart. Highly similar images are difficult to identify in matching-to-sample tests. Memory tests after long delays are more difficult than those following short delays. Stimulus magnitude, image similarity, and delay interval are all types of publicly available information that indicate the difficulty of a particular test trial. Subjects performing tests with such stimuli might use the identity, magnitude, similarity, delay, or other publicly available information as a discriminative cue for declining tests or rating confidence. For example, if subjects have experienced low rates of reward with stimuli in a specific magnitude range, they could learn to avoid tests with all stimuli in that range (see Kornell, et al., 2007; S. J. Shettleworth & Sutton, 2003 for the same argument). In a somewhat more subtle version of this account, extra-experimental events that might interfere with attention or performance (e.g. randomly occurring noises in the test environment, itches, or bouts of auto-grooming) can become discriminative stimuli for the metacognitive response (e.g. Hampton, 2001, 2005). The probability that *Environmental Cue Associations* can account for performance in a given paradigm is best assessed by generalization tests which determine whether or not performance is maintained across changes in the particular stimuli used and specific conditions of testing. If performance immediately generalizes to new test conditions or new stimuli, it is safe to conclude that metacognitive responding was not controlled by stimuli that were changed for the generalization test.

Behavioral cue associations

This account of metacognitive behavior is similar to *Environmental Cue Associations*, with the exception that the discriminative stimuli controlling use of the metacognitive response are systematically generated by the subject in a way that correlates with accuracy in the primary task. For example, the subject may vacillate when it does not know the correct response on a given test (Muenzinger, 1938; Tolman, 1948). This vacillation does not necessarily represent metacognition by the subject that it does not know the answer, but can rather be an unmediated result of not knowing how to respond. It is common to see this sort of vacillation in monkeys taking matching-to-sample tests, for example, in which they look back and forth between the choice stimuli before choosing (personal observations). It is also well known that response latency is often longer for incorrect than correct responses. Because vacillation and response latency correlate with accuracy, subjects could use these self-generated cues as discriminative stimuli for the metacognitive response, for example by declining tests on which they experience a relatively long response latency. One way to assess whether *Behavioral Cue Associations* account for metacognitive performance is to require subjects to make the secondary metacognitive judgment before they have seen the relevant primary test, and therefore before the test could have elicited vacillation or similar behavioral responses (Foote & Crystal, 2007; Hampton, 2001).

Response competition

In most reports of metacognition in nonhumans, subjects are confronted with the primary discrimination problem or memory test and the secondary metacognitive response option simultaneously (e.g. Basile, Hampton, Suomi, & Murray, 2009; Call & Carpenter, 2001; Hampton, Zivin, & Murray, 2004; Expt. 1 in Inman & Shettleworth, 1999; Shields, et al., 1997; J. D. Smith, et al., 1995; J.D. Smith, et al., 1998; Washburn, Smith, & Shields, 2006). Because subjects can only make one response (a primary test response or a secondary decline test response, for example), simultaneous presentation puts these two behaviors in direct competition. As indicated above, animals are frequently slower to respond on error trials than on correct trials. On error trials with no prepotent primary test response, the probability that the subject will make the secondary metacognitive decline test response is greater, simply because no other competing response occurs immediately. On correct trials, when the inclination to make a primary test response is strong, it may dominate the tendency to decline the test or collect more information before responding. In all of the studies cited above, the evidence for metacognition is that difficult primary test trials are declined or delayed (while more information is collected). The higher probability of the metacognitive response on difficult trials may therefore result from competition between primary choice responses and secondary metacognitive responses. For an example of how different behaviors can compete, consider a rat that has good knowledge of the location of food on a maze. Such a rat is likely to go directly to the baited locations and is consequently unlikely to explore other locations or engage in other behavior. *Response Competition* can be ruled out as an account for metacognitive responding by presenting the secondary metacognitive response option either *before or after* the primary test, so that the two types of response do not compete directly.

Introspection

Metacognition could also be mediated by private introspective assessment of the subject’s mental states. While introspection (i.e. the contemplation or perception of one’s own mental states) might not necessarily require consciousness, it is closely allied with consciousness in humans (Koriat, 1996; Nelson, 1996). By the introspection account, the discriminative stimulus controlling a metacognitive response (e.g. declining to take a test) is the private experience of uncertainty (J.D. Smith, Shields, & Washburn, 2003) or the weakness of memory (Hampton, 2001, 2005). In the case of uncertainty, subjects are suggested to experience conscious (at least in humans) “feelings of uncertainty” that differ from the experience of objective stimuli (J.D. Smith, et al., 2003). In the case of memory, subjects are proposed to assess the strength of their memory. The assessment of memory might be accomplished through several mechanisms that vary in sophistication from detecting whether a memory is present (while knowing nothing of the content of the memory) to attempting to retrieve the relevant memory and determining the success of that effort (Figure 1). Subjects use the decline response or other metacognitive response when memory is determined to be absent or weak (Hampton, 2005, 2006). The important difference between this account and the preceding three is that use of the metacognitive response is based on privileged introspective access to the subject’s cognitive states, rather than on publicly available information or *Response Competition*. Due to the private nature of *Introspection*, the conclusion that it accounts for metacognitive performance in nonhumans can probably be reached only by ruling out other accounts.

Evaluation of the literature through selected examples

It is exciting that in the time since publication of the first paper specifically addressing the question of nonhuman metacognition this literature has grown to the point that it is not feasible to comprehensively review it in this short article. Instead, I will evaluate a set of representative studies with respect to the four mechanisms for metacognition described above. If I have omitted a reader’s favorite study, I apologize and hope that the current analysis can be readily extended to additional cases. A summary of this selected analysis of the literature is provided in Table 1.

Dolphin Auditory Psychophysics (J. D. Smith, et al., 1995)

The first report of metacognition in a nunhuman species described the performance of a bottlenosed dophin (*Tursiops truncatus*) in an auditory psychophysical task (J. D. Smith, et al., 1995). This publication nicely introduces many of the major features common to tests of nonhuman metacognition. The paradigm is also relatively straightforward compared to more elaborate designs that followed. For these reasons, I will use this example to illustrate much of the current approach to analyzing nonhuman metacognition findings, referring back to it in later sections. Analysis of this pioneering publication will not, therefore, take into account subsequent data and analyses that followed from the same research group, but will instead illustrate key difficulties in determining how metacognitive responses are controlled in nonhuman species.

A dolphin was required to discriminate between tones of 2100-Hz and tones of any lower frequency (ranging from 1200–2099 Hz). The dolphin was initially trained to make this primary discrimination by responding to a left paddle following 2100 Hz tones and to a right paddle for any lower frequency tone. As expected, the dolphin's accuracy decreased as the tested frequency approached 2100 Hz (the dolphin was likely to respond to the left paddle when the frequency was close to 2100 Hz, treating these tones as if they were 2100 Hz tones). After the dolphin had acquired this primary discrimination, a third paddle was introduced that allowed the dolphin to decline a given discrimination trial in favor of an easy discrimination (a 1200 Hz tone). With these contingencies in place, the dolphin could maximize the rate of reward by performing the primary discrimination (choosing the left or right paddle) when the discrimination was easy while selecting the third paddle when the discrimination was difficult. The dolphin's behavior generally conformed to these contingencies. The dolphin was unlikely to use the third paddle following low frequencies (the easiest trials) and was increasingly likely to use this "decline test" paddle following frequencies near 2100 Hz (the most difficult trials).

The dolphin clearly met the criteria for metacognition, adaptively taking easy tests and declining difficult tests. How might the dolphin have accomplished this? Several features of this experiment suggest that the dolphin may have used publicly observable cues to guide use of the decline response, suggesting public metacognition. First, the dolphin may have used tone frequency as a discriminative stimulus for making a decline response to the third paddle. In this design, discrimination difficulty is confounded with frequency, that is, difficulty and frequency are correlated. The dolphin may have learned to select the decline test paddle in the presence of stimuli belonging to a particular frequency range because of its reinforcement history with those particular tones, rather than because of a subjective feeling of uncertainty. Thus, *Environmental Cue Associations* may be sufficient to account for the metacognitive performance. This account could be tested by determining whether use of the third paddle response generalized immediately to tests conducted in new frequency ranges (e.g., 3100 Hz vs. 2200–3099 Hz). If the dolphin had learned a general metacognitive response, it should continue to avoid difficult trials in the new frequency range. By contrast, if the dolphin had learned to use the decline test response whenever tones of specific stimuli were used, the dolphin would have to relearn which frequencies should occasion this response through trial and error learning of which frequencies were associated with low rates of reward. Second, the dolphin may have used its own publicly observable behavior as a discriminative stimulus for declining tests. As described earlier, subjects often vacillate on difficult trials, a pattern also reported for the dolphin as "ancillary behaviors" near threshold (J. D. Smith, et al., 1995). It is tempting to interpret these "ancillary behaviors" metacognitively, as indicating that the subject hesitates because it knows it is uncertain. However, it is safer to interpret them non-metacognitively; when the subject does not know how to respond, it is slow to do so and may engage in other behavior in the meantime. Thus, the dolphin may have learned to use its own vacillation as a discriminative stimulus for the decline response, a type of *Behavioral Cue Association.* Third, in this experiment (and many others) the secondary metacognitive response and the primary choice response were presented simultaneously, admitting the possibility that *Response Competition* can account for metacognitive performance. Simultaneous presentation places selection of one of the primary test responses (left or right paddle) in direct conflict with selection of the secondary decline response (the third paddle). On difficult trials with no clear correct response, the tendency to respond to the left or right paddle is low. Simply because it reduces the probability of a left or right response, difficulty in the primary task may increase the probability of the decline test response. Finally, the dolphin may have used *Introspection*, or private metacognition. By this account, the dolphin reacted to a private cognitive discriminative stimulus (e.g. subjective uncertainty) that indicated that it did not know the correct answer on specific trials. Because multiple public accounts are viable, invoking an introspective account may be unwarranted. Note that the same critique applies to the similar studies of monkeys performing psychophysical pixel density tasks (Shields, et al., 1997).

Collecting Information When Ignorant (Basile, et al., 2009; Call & Carpenter, 2001; Hampton, et al., 2004)

Metacognition is evident when subjects collect additional information when ignorant and act immediately when informed. Call and Carpenter (2001) developed a clever test of this capacity and used it with human children, chimpanzees (*Pan troglodytes*), and orangutans (*Pongo pygmaeus*). A modified version of this same test was subsequently used with rhesus monkeys (*Macaca mulatta*, Hampton, et al., 2004) and capuchin monkeys (*Cebus apella*, Basile, et al., 2009). Subjects were presented with a set of opaque tubes in which food was hidden. Subjects either witnessed the baiting (seen trials) or did not (unseen trials), and therefore were either informed or ignorant about the food's location on each trial. At test, subjects could select a single tube and collect the reward, if correct. This test is an interesting assessment of metacognition because the subjects could bend over and look down the length of the tubes to locate the food before choosing (see Figure 2). Subjects demonstrate metacognition by collecting information when ignorant (unseen trials) and choosing immediately when informed (seen trials). Human children, chimpanzees, orangutans, and rhesus monkeys clearly showed this pattern of behavior, while the case for capuchin monkeys was less clear (some capuchins made this differentiation under at least some conditions).

How does this performance relate to the four accounts of metacognitive behavior under consideration? This discussion will focus on the representative study with rhesus monkey subjects (Hampton, et al., 2004). Unlike with the case of the dolphin psychophysical test, it is not possible that the rhesus used the identity of the test stimuli to guide their decision to look because each tube was equally likely to contain the food on both seen and unseen trials. Monkeys were familiarized with the apparatus and procedures (including gaining experience with looking down tubes) in such a way as to prevent them from learning via differential reinforcement to look selectively on unseen trials. Furthermore, comparatively few critical test trials were presented, to prevent monkeys from developing associations between experimental cues and the probability of reinforcement (see Hampton et al., 2004 for further details of how this was accomplished). It is therefore unlikely that *Environmental Cue Associations* or *Behavioral Cue Associations* underlie performance in these tests. These tests, like many other tests of metacognition, presented the primary choice response and the secondary metacognitive response simultaneously. That is, the opportunity to look down the tubes partly overlapped in time with the opportunity to choose a tube (although monkeys had 2 seconds of opportunity to look through a clear screen before they could actually select a tube). Knowing the food's location may strongly predispose a monkey to select that tube, decreasing the occurrence of all other possible behaviors, including searching the tubes. Consequently, the metacognitive performance of monkeys in this paradigm may be the result of *Response Competition*. Finally, monkeys may have used *Introspection* about their own knowledge state to determine whether they needed to look before selecting a tube. However, as the behavior can be explained by at least one public mechanism (*Response Competition*), more research is needed before we can safely infer a private mechanism.

Serial Position and Confidence about Memory (J.D. Smith, et al., 1998)

When subjects are presented with lists of items to remember (such as the list of salad dressings available with your order at a restaurant), it is typical for items early and late in the list to be remembered better than middle items. Such serial position effects have been a staple of memory research in humans and nonhumans (e.g. Wright, Santiago, Sands, Kendrick, & Cook, 1985). Smith et al. (1998) took advantage of this predictable pattern of memory strength to assess whether monkeys showed metacognition for memory in a serial probe recognition task. Monkeys saw a list of four consecutive random dot polygon figures and their memory for individual polygons from the list was probed using a yes-no recognition test. Monkeys showed the expected serial position effect; their memory was better for the first and last items than for the middle items. Monkeys were then presented with a decline test response, similar to that used with the dolphin psychophysical test (Smith et al., 1995), concurrently with a probe polygon that may or may not have been from the studied list. Consistent with adaptive metacognitive control, the monkeys declined tests of the middle list items more often than tests of the first and last list items.

How does the monkeys' performance in this paradigm relate to the four accounts of metacognitive behavior under consideration? Because 101 different polygons were randomly assigned to the four list positions, the identity of the images (even if some were easier to remember than others) would not correlate with the list position effect. Furthermore, single polygons were presented for recognition at test, with no indication of which list position they occupied during study. Because no publicly observable aspect of the test setup correlated with memory strength, it is unlikely that the metacognitive response was under control by *Environmental Cue Associations*.

As indicated earlier, longer response times are typically associated with incorrect responses, so response latency could serve a as a discriminative stimulus for the decline response. Because the monkeys had ample opportunities during training to associate long response latencies with sparser rewards following the primary choice response, metacognitive performance could be under the control of *Behavioral Cue Associations*. Evaluation of an unusual pattern in the use of the decline response in this study may help determine the likelihood of this possibility. On each trial of this yes-no recognition test, the primary response was to classify the probe stimulus as either from the just studied list (yes) or not (no). Both monkeys tested were most likely to use the decline test response on trials in which the probe stimulus was not from the studied list (J.D. Smith, et al., 1998, Figure 7, A and B, "NT" responses). For monkey Abel, this frequent use of the decline response tracks his relatively poor performance on these trials (i.e. his metacognitive choices were accurate; he declined the difficult "NT" trials). For monkey Baker, however, accuracy is quite high on these "NT" trials and should be associated with infrequent use of the decline test response (i.e. his metacognitive choices were inaccurate; he declined easy trials). This seeming incongruity could possibly be addressed through detailed examination of their response latencies. Did Baker take a relatively long time to report that items were not from the studied list, while Able did so relatively rapidly? If so, this would support the argument that response latency controlled use of the decline response in this paradigm, even when it did not correlate with accuracy.

Like all examples discussed thus far, the primary yes-no recognition test and the secondary decline response were presented simultaneously. Thus, *Response Competition* could account for the observed metacognitive performance. Finally, *Introspection* also remains a viable account of metacognitive performance in this study. However, we must again be cautious in inferring *Introspection* until we can rule out possible public mechanisms.

Retrospective Metacognitive Judgments

Probably the most creative of the published nonhuman metacognition paradigms is the retrospective gambling paradigm (Kornell, et al., 2007; Son & Kornell, 2005). In this paradigm, monkeys rated their "confidence" by wagering either a large or small number of video tokens on the accuracy of each test trial immediately after they completed it. The video tokens were secondary reinforcers that were periodically "cashed out" for actual food when a sufficient number had accumulated. Critically, monkeys placed their wager after answering, but before receiving feedback about their accuracy. In this paradigm, metacognition predicts large wagers following easy tests (i.e. when monkeys are confident of their answer) and small wagers following difficult test (i.e. when monkeys would be unsure of their answer). This is indeed how the monkeys performed, suggesting that they knew whether they had responded correctly despite the lack of feedback prior to placing their bet.

Presentation of the metacognitive response after completion of test trials effectively rules out *Response Competition* as a viable account for metacognitive performance; that is, performing the primary test response does not directly lower the probability of performing the secondary metacognitive response. Kornell et al. (2007) also ruled out *Environmental Cue Associations* by showing that use of the metacognitive gambling response generalized across stimuli and, more importantly, across test types (from perceptual tests to a mnemonic test).

Behavioral Cue Associations remain a potential source of metacognitive control in these studies. Although separating the secondary metacognitive response from the primary task is a powerful control procedure, offering the metacognitive response *after* the primary test means that the subjects have already directly experienced the difficulty of each trial before they have to make their wager. Behavioral cues such as response latency are therefore available as discriminative stimuli to control the subsequent metacognitive response. Indeed, Kornell et al. (2007) report that longer response latencies were associated with both incorrect responses and small bets (which indicate low confidence). Unfortunately, transfer from perceptual to mnemonic tasks does not rule out control by response latency; the same association of long response latency with difficult trials is likely maintained across tasks and provides a basis for generalization of the metacognitive response. Finally, *Introspection* also remains a viable basis for control of the metacognitive response in these studies. Future studies might focus on ruling out stimulus control by response latency.

Prospective Metacognitive Judgments

A few studies have required subjects to make a metacognitive judgment *before* seeing the actual test (Foote & Crystal, 2007; Hampton, 2001; Inman & Shettleworth, 1999; Suda-King, 2008; Sutton & Shettleworth, 2008). Presentation of the secondary metacognitive choice prior to the primary test choice has at least two positive attributes. First, as described above, *Response Competition* cannot account for metacognitive performance because the primary and secondary responses are not available simultaneously. Second, *Behavioral Cue Associations* cannot account for metacognitive performance because the subject has not yet seen the test when offered the metacognitive response and, therefore, cannot use vacillation or "ancillary responses" as discriminative stimuli for the metacognitive response.

Unfortunately, presenting the secondary metacognitive response before the primary test response does not by itself rule out all public mechanisms. While it is the case that subjects in the Suda-King (2008) study chose whether to decline tests prior to presentation of the very final test, the test stimuli were plainly visible at the time subjects made the metacognitive response. Due to this failure to fully separate the metacognitive response from presentation of the test stimuli, it is not clear that this arrangement represents a true prospective metacognitive judgment. The otherwise elegant study of metacognition in a temporal psychophysical task by

Foote & Crystal (2007) did not include a generalization test involving new discrminanda. Without a generalization test it is not possible to determine the extent to which the use of the decline test response is tied to specific test stimuli. Thus, it is possible that, despite the separated primary and secondary responses, the metacognitive response found by Foote & Crystal was controlled by *Environmental Cue Associations*. The two papers describing using this technique with pigeons did not report metacognitive performance. For these reasons, the remaining discussion of prospective metacognitive judgments will focus on the study of metamemory in rhesus monkeys (Hampton, 2001).

In the Hampton (2001) study, monkeys were initially trained to match to sample, and then the delay between the study and test phases was gradually lengthened until monkeys performed at an intermediate level between chance and perfection. A metacognitive response was then introduced at the end of the delay interval that allowed monkeys to accept the memory test and receive a favored reward if correct, or decline the memory test and receive a guaranteed, but less desirable, reward. On other trials, only the option to take the memory test was offered at the end of the delay. Monkeys were more accurate on trials on which they accepted the test than on trials on which they were required to take the test, demonstrating that they accepted tests when memory was relatively good and declined tests when memory was relatively poor. Use of the decline response generalized to conditions in which memory was directly manipulated either by providing no sample to remember (monkeys overwhelming declined subsequent memory tests) or by increasing the delay interval (monkeys were more likely to decline tests after long than after short delay intervals).

How does the monkeys' performance in this paradigm relate to the four accounts of metacognitive behavior under consideration? *Environmental Cue Associations* cannot control the metacognitive response because monkeys generalized to new test stimuli every day, to trials with no sample to be remembered, and to trials with different delay intervals. *Behavioral Cue Associations* are also unlikely to control the metacognitive response because monkeys had to choose to accept or decline tests before they had seen the test and, thus, before they could exhibit these behaviors. A narrow exception is that it is possible that there are two or more "behavioral states" that the monkeys could be in, one that promotes attention and memory and another that does not. One can imagine a situation where the monkeys can recognize their current state as attentive or inattentive (possibly based on body posture or general arousal) and differentially associate each state with an adaptive response (accepting and declining tests, respectively). However, because the monkeys generalized use of the decline response to no-sample trials, such an explanation would require the additional property that the "attentive state" is triggered by presentation of a sample.

Because the secondary metacognitive response was made before the test stimuli were presented, and it was not in direct competition with the primary test responses, *Response Competition* cannot account for metacognitive performance in this paradigm. By process of elimination, *Introspection* appears to be the most likely candidate for control of the metacognitive response. However, this conclusion should be made only tentatively given the diversity of possible alternative explanations, only a subset of which have been considered here.

Converging evidence?

I have organized this article around a broad functional definition of metacognition that emphasizes achieving adaptive cognitive control by any mechanism possible. In a practical context, such a definition is entirely sensible. For example, we may be satisfied that our students learn to effectively regulate their study habits, regardless of whether this regulation is based on publicly observable cues, such as the quantity of material to be learned, or on introspections, such as repeatedly attempting to recall studied material. In this broad context, the literature I have reviewed certainly provides converging evidence for metacognition in nonhumans. The reviewed studies show that animals adaptively regulate decisions about when to take tests, when to collect more information, and how to rate their own performance. It is much less clear whether these studies provide converging evidence regarding the mechanisms by which this adaptive cognitive control is achieved. Inspection of Table 1 shows that *Introspection* is always a potential source of stimulus control in these studies; however, even consideration of this limited set of possible alternative accounts shows that metacognitive control in most studies can be adequately explained without invoking introspection. Thus, there is a high bar to clear in terms of ruling out alternative mechanisms for metacognition before we can conclude that any nonhuman animals engage in private metacognition.

What are we trying to learn? Relationship to implicit and explicit representation

Some investigators may want to limit use of the term metacognition more strictly than I have done here. In particular, metacognition is often associated with conscious awareness and introspection; many investigators might argue that introspective metacognition is the most interesting case (Nelson, 1996). The rationale for a more restrictive definition might parallel analyses of the early memory experiments of Hunter, using delayed response (see S.J. Shettleworth, 1998 p. 239–242). In his experiments subjects were restrained in a start box from which they saw a cue light illuminate and extinguish above one of several doors, indicating the location of a food reward. During a delay interval, subjects had to remember which door to approach. One way animals “solved” this test was to remain oriented toward the correct door during the delay interval. Using this method, the subjects did indeed “remember” and responded correctly, but such performance did not require a mental representation of the cued location that would persist even if the animal moved or its view of the apparatus was occluded. To most investigators of learning and memory, mental representations are of considerably more interest than successful postural mediation. Similarly, most researchers of nonhuman metacognition will be more interested in paradigms that rule out adaptive control by public mechanisms.

If the study of metacognition is motivated mostly by the possibility that it provides a means for studying something akin to conscious introspection in animals, then we need to be thorough in our use of procedures that rule out other sources of stimulus control. Studies of metamemory, in particular, are aimed at determining whether we can make a distinction between implicit and explicit mental representations in nonhuman species that parallel those made in humans (Hampton, 2001, 2003, 2005, 2006; Hampton & Hampstead, 2006; Hampton, et al., 2004). Perhaps the first studies to address explicit representation in nonhumans were the “blindsight” studies done in monkeys (Cowey & Stoerig, 1995). These studies showed that monkeys can accurately localize a stimulus even when they report that no stimulus is present in a present-absent discrimination. Subsequently, similar techniques were used in experiments that assessed metacognitive abilities. These demonstrations depend on the capacity of subjects to make what Weiskrantz (2001) called a “commentary response,” which is interpreted to reflect some assessment by subjects of their subjective perceptual experience. It may still be premature to conclude that any case of observed metacognition in nonhumans depends on introspection involving explicit representations, but when sources of public stimulus control are eliminated, it is more likely that *Introspection* underlies metacognitive performance.

Implications for Comparative Psychology

It is intriguing that it appears to be easier to demonstrate metacognition in some species than in others. For example, while there are many reports of metacognition in rhesus monkeys (e.g.

Hampton, 2001; Kornell, et al., 2007; J.D. Smith, et al., 2003), work with pigeons has been much less likely to detect metacognition (Inman & Shettleworth, 1999; Sole, et al., 2003; Sutton & Shettleworth, 2008). In parallel tests conducted with human children, apes, rhesus monkeys, and capuchin monkeys, capuchin monkeys show by far the weakest evidence for metacognition (Basile, et al., 2009; Call & Carpenter, 2001; Hampton, et al., 2004). It is tempting to interpret these differences as indicating that metacognitive control is not "easy," is unlikely to come about through "simple" associative learning (of which pigeons and capuchins are certainly capable), and may be restricted to relatively few species. However, it is still too early to reach this conclusion; there are a host of species characteristics that may interfere with performance in metacognitive tests (e.g. differences in attention, impulse control, and motivation). Sorting out which species can and cannot behave metacognitively will be greatly helped if we can agree as a community what behavioral criteria are required for 1) metacognition, and 2) introspective control of metacognition. We then need to specifically design experiments to evaluate performance with respect to these criteria. Hopefully the ideas put forward here will contribute to developing these new criteria and new designs.

Acknowledgments

Ben Basile, Victoria Templer, and Emily Brown provided comments on an earlier version of this manuscript. Preparation of this manuscript was supported by NIH grant 1RO1MH082819-01A1, by Yerkes Center base grant No. RR-00165 awarded by the Animal Resources Program of the National Institutes of Health, and by the Center for Behavioral Neuroscience under the STC Program of the National Science Foundation under Agreement No. IBN-9876754.

References

Basile BM, Hampton RR, Suomi S, Murray EA. An assessment of memory awareness in tufted capuchin monkeys (*Cebus apella*). Animal Cognition 2009;12:169–180. [PubMed: 18712532]

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

Clark RE, Squire LR. Classical conditioning and brain systems: The role of awareness. Science 1998;280:77–81. [PubMed: 9525860]

Cohen NJ, Eichenbaum H, Deacedo BS, Corkin S. Different memory systems underlying acquisition of procedural and declarative knowledge. Annals of the New York Academy of Sciences 1985;444:54–71. [PubMed: 3860122]

Cowey A, Stoerig P. Blindsight in Monkeys. Nature 1995;373:247–249. [PubMed: 7816139]

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]

Hamann SB, Squire LR. Intact perceptual memory in the absence of conscious memory. Behavioral Neuroscience 1997;111:850–854. [PubMed: 9267663]

Hampton RR. Rhesus monkeys know when they remember. Proceedings of the National Academy of Sciences of the United States of America 2001;98:5359–5362. [PubMed: 11274360]

Hampton RR. Metacognition as evidence for explicit representation in nonhumans. Behavioral and Brain Sciences 2003;26:346–347.

Hampton, RR. Can Rhesus monkeys discriminate between remembering and forgetting?. In: Terrace, HS.; Metcalfe, J., editors. The missing link in cognition: Origins of self-reflective consciousness. New York: Oxford University Press; 2005. p. 272-295.

Hampton, RR. Memory awareness in rhesus monkeys. In: Fujita, K.; Shoji, I., editors. Diversity of Cognition. Kyoto: Kyoto University Press; 2006. p. 282-299.

Hampton RR, Hampstead BM. Spontaneous behavior of a rhesus monkey (*Macaca mulatta*) during memory tests suggests memory awareness. Behavioral Processes 2006;72:184–189.

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;7:239–254. [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.

Knowlton BJ, Ramus SJ, Squire LR. Intact Artificial Grammar Learning in Amnesia - Dissociation of Classification Learning and Explicit Memory for Specific Instances. Psychological Science 1992;3:172–179.

Knowlton BJ, Squire LR. The Learning of Categories - Parallel Brain Systems for Item Memory and Category Knowledge. Science 1993;262:1747–1749. [PubMed: 8259522]

Koriat, A. Memory's knowledge of its own knowledge: The accessibility account of the feeling of knowing. In: Metcalfe, J.; Shimamura, AP., editors. Metacognition. Cambridge, MA: The MIT Press; 1996. p. 1-25.

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

Muenzinger KF. Vicarious trial and error at a point of choice: A general survey of its relation to learning efficiency. Journal of Genetic Psychology 1938;53:75–86.

Nelson TO. Consciousness and metacognition. American Psychologist 1996;51:102–116.

Nelson TO, Narens L. Metamemory: A theoretical framework and new findings. Psychology of Learning and Motivation 1990;26:125–322.

Roberts, WA. Principles of Animal Cognition. Boston: McGraw Hill; 1998.

Shettleworth, SJ. Cognition, Evolution, and Behavior. New York: Oxford University Press; 1998.

Shettleworth SJ, Sutton JE. Animal metacognition? It's all in the methods. Behavioral and Brain Sciences 2003;26:353–354.

Shields WE, Smith JD, Washburn DA. Uncertain responses 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, Schull J, Strote J, McGee K, Egnor R, Erb L. The Uncertain Response in the Bottle-Nosed-Dolphin (*Tursiops-Truncatus*). Journal of Experimental Psychology-General 1995;124:391–408. [PubMed: 8530911]

Smith JD, Shields WE, Washburn DA. Memory monitoring by animals and humans. Journal of Experimental Psychology-General 1998;127:227–250. [PubMed: 9742715]

Smith JD, Shields WE, Washburn DA. The comparative psychology of uncertainty monitoring and metacognition. Behavioral and Brain Sciences 2003;26:317–374. [PubMed: 14968691]

Sole LM, Shettleworth SJ, Bennett PJ. Uncertainty in pigeons. Psychonomic Bulletin & Review 2003;10:738–745. [PubMed: 14620372]

Son, LK.; Kornell, N. Metaconfidence judgments in rhesus macaques: Explicit versus Implicit mechanisms. In: Terrace, HS.; Metcalfe, J., editors. The missing link in cognition: Origins of self-reflective consciousness. Oxford: Oxford University Press; 2005.

Suda-King C. Do orangutans (*Pongo pygmaeus*) know when they do not remember? Animal Cognition 2008;11:21–42. [PubMed: 17437141]

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]

Tolman EC. Cognitive maps in rats and men. The Psychological Review 1948;55:189–208.

Tulving E, Schacter DL. Priming and Human-Memory Systems. Science 1990;247:301–306. [PubMed: 2296719]

Washburn DA, Smith JD, Shields WE. Rhesus monkeys (*Macaca mulatta*) immediately generalize the uncertain response. Journal of Experimental Psychology- Animal Behavior Processes 2006;32:185–189. [PubMed: 16634662]

Weiskrantz L. Commentary responses and conscious awareness in humans: The implications for awareness in non-human animals. Animal Welfare 2001;10:S41–S46.

Wright AA, Santiago HC, Sands SF, Kendrick DF, Cook RG. Memory Processing of Serial Lists by Pigeons, Monkeys, and People. Science 1985;229:287–289. [PubMed: 9304205]

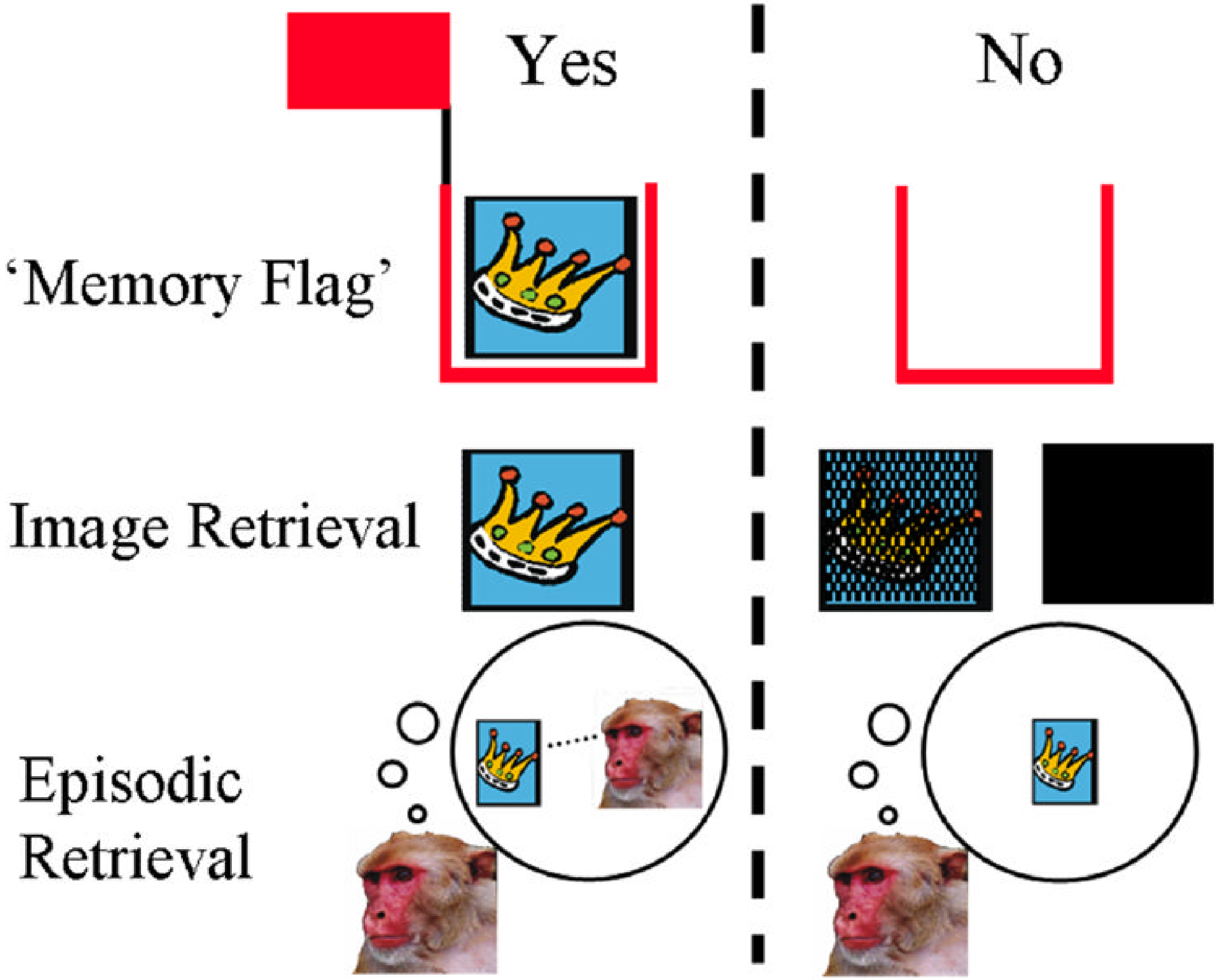

Figure 1.

Cartoons of three candidate processes of introspective memory assessment. The column to the left of the dashed line represents the contents of cognitive processing on trials on which monkeys choose to take the memory test. The right column depicts the same on trials on which the test is declined. The memory flag hypothesis posits an indicator for the presence of memory. Monkeys use the metacognitive response contingent on the indicator, but are not aware of the content of the memory. In the case of the image retrieval model, the decision to take the memory test is based on the vividness of the memory retrieved. The episodic retrieval hypothesis proposes that monkeys take the test when they can remember the context of the study episode and decline the test when this information cannot be recovered.

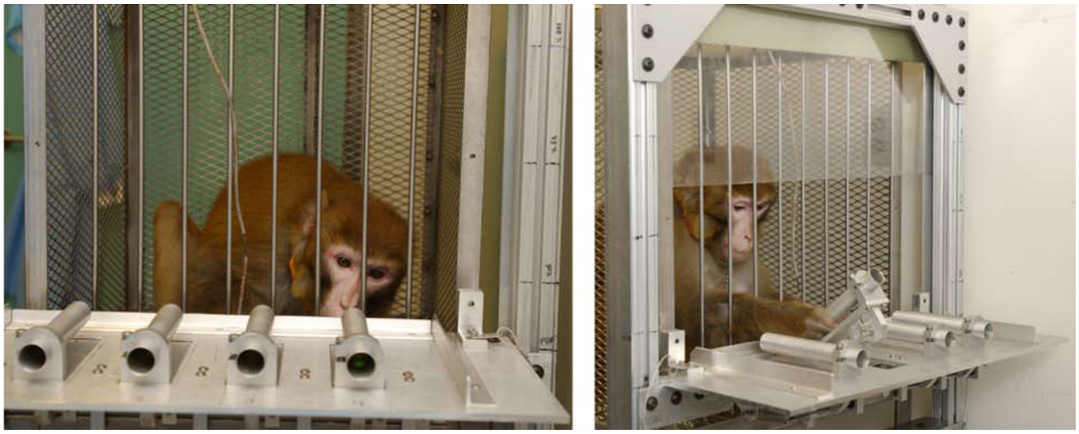

Figure 2.
Left, a rhesus monkey, ignorant of the food's location, collects more information before making a choice. Right, an informed monkey makes a choice without going to the effort of confirming the location of the food. Such selective information seeking suggests that the monkey knows when he knows, and only seeks more information as needed.

Table 1

Characterization of selected experiments with respect to four classes of stimulus control for metacognitive responding. A green background indicates that the type of stimulus control indicated in the column heading can account for the reported metacognitive performance. A medium red background indicates that the indicated stimulus control is ruled out. A light yellow background indicates a low probability of stimulus control. Text indicates how a particular type of stimulus control was ruled out. To the extent that particular sources of stimulus control can be ruled out, the remaining sources of control are more likely to be in effect.

	Distinctive features	Environmental cues	Behavioral cues	Response competition	Introspection
Call & Capenter, 2001; Hampton, Zivin & Murray, 2004	opaque tubes, “spontaneous” metacognition	unlikely, counter-balanced stimuli	unlikely, limited experience	possible, concurrent responses	possible
Foote and Crystal, 2007	prospective tests, temporal psychophysics	possible	unlikely, prospective judgment	no, prospective judgment	possible
Hampton, 2001	prospective judgment, generalization tests	no, generalization to no-sample trials, delays	unlikely, prospective judgment, generalization to no-sample trials	no, prospective judgment	possible
Kornell, Son & Terrace, 2006	retrospective confidence judgment by “gambling”	no, generalized to new stimuli	possible, response latency	no, retrospective judgment	possible
Smith, Schull, Strote, McGee, Egnor & Erb, 1995	dolphin auditory psychophysics	possible	possible	possible	possible
Smith, Shields, Schull & Washburn, 1997	psychophysical pixel density test	possible	possible	possible	possible
Smith, Shields, Washburn & Allendoerfer, 1998	serial position effect	no, list position not perceptible at test	possible	possible	possible