Speaking about Metacognition:
A Developing Ability in Young Children

Richard A. Nathanson
Haverford College
Abstract

The development of metacognition in nursery school children was examined in this study. Our computer program encouraged the acquisition of a perceptual discrimination based on the density of dots in a rectangle. Children were rewarded with small, attractive stickers on correct responses, and were delayed from winning more stickers on incorrect responses. Once they attained the discrimination, the program made available a third response, one that provided an escape to the discrimination task and gave automatic rewards after a short delay. The most suitable explanation for escaping from a trial is that the subject is responding to a subjective internal state; he knows that he doesn’t know. We also asked subjects to verbally describe their strategies for responding to the discrimination and for when they choose to escape from the discrimination. We found that the ability to verbalize about one’s own metacognition is a fundamentally different and higher-order ability than is the ability to behave metacognitively. Even when children are capable of metacognition, they don’t verbalize about that ability, suggesting that nonverbal stimuli and assessments are most appropriate for studying metacognition in children.
Speaking about Metacognition:
A Developing Ability in Young Children

Metacognition, "any knowledge or cognitive activity that takes as its object, or regulates, any aspect of any cognitive enterprise" (Flavell, 1985, p. 104), is an important part of psychological research because it is central to many philosophical and psychological questions about human thought. It is a broad field, in part because it has been discussed from many different intellectual and experimental perspectives. In this paper, I will first review the current literature defining metacognition in order to focus its definition as it applies to this study and as it fits into the conceptual hierarchy of metacognitive behaviors based on consciousness. Second, I will review in more detail Smith, Schull, and Washburn’s (in preparation) work, as it forms the basis for our experiment. Third, I will consider language as a precursor for metacognition. Fourth, I will review work on metamemory and metacomprehension both as it is based in verbal technique and as it departs from a dependence on verbal measures as a background for the development of this task. Finally, I will summarize the reasoning behind this experiment and present the study questions.

A Conceptual Hierarchy of Metacognition

The definition of metacognition is not consistent in the psychological and philosophical literature to date; the term is used to refer to a broad range of abilities. In this study, however, we seek to demonstrate that metacognition develops in human children by focusing on a specific type of metacognition and giving it a specific operational definition: the ability to unambiguously and behaviorally signal a perception of internal, subjective uncertainty about a discrimination task. To that
end, it is important to have an understanding of the framework within which the specific phenomenon in question falls. By placing our task in this hierarchy, I will clearly define what we mean by ‘metacognition’ in this study; I make no claims about demonstrating more conscious, higher-order metacognition.

No consciousness is required at the first level in this metacognitive hierarchy of increasing self-awareness and self-regulation. Flavell (1985) terms behavior at this level a simple “metacognitive experience” (p.110); cognition about cognition takes place, but the actor is not self-reflectively aware that he is thinking about his own thoughts. Our metacognition falls here, at the beginning of a metacognitive hierarchy as constructed by the amount of consciousness required.

The second level of metacognitive functioning involves a more conscious awareness that one’s activities are metacognitive. This “metacognitive knowledge” (Flavell, 1985, p. 110) is concerned with the ability to understand that one’s activities are metacognitive. Metacognitive knowledge requires more consciousness than a metacognitive experience because it demands an ability to look reflectively at one’s own thoughts. In an experimental situation, this translates into “the ability to mention, or consciously describe and discuss, one’s own cognitive activities” (Campione et al., 1982, p. 457); self-report techniques are employed to encourage the subject to become conscious of, and then report, his own metacognitive knowledge.

Metacognitive control comprises the next level of metacognitive functioning. Campione et al. (1982) refer to this level as “executive control” and include “self-

1“Passive knowledge” (Brown, Bransford, Ferrara, and Campione, 1982), and “reflective access” (Campione, Brown, and Ferrara, 1982), while slightly different, also refer to abilities at this level.
regulatory mechanisms ... during an ... attempt to solve problems” (p. 434). In comparison to reflective access in metacognitive knowledge, metacognitive control involves multiple access, the “flexible use of information available to the system,” including the abilities to check, plan, monitor, test, revise, and evaluate strategies concerning a cognitive task while performing that task (p. 457). At this level, behavior is reflectively and consciously controlled using metacognitive insights gained through multiple access. Piaget further divides metacognitive executive control into three categories which form a sub-hierarchy of increasing cognitive ability: autonomous, active, and conscious (in Brown, et al., p. 117-119). At the autonomous level, learners simply regulate and change their actions. The active level reflects Karmiloff-Smith and Inhelder’s notion of a “theory-in-action” (in Brown, et al., p. 118): learners construct and test theories in a concrete fashion. Finally, in Piaget’s conscious regulation, hypotheses are tested in an abstract mental way, without reliance on concrete examples. Metacognitive control is metacognitive knowledge applied adaptively and consciously to a situation.

Epistemic cognition is the final step in the hierarchy of metacognition. It requires sophisticated conscious self-reflection. Kitchener (1983) defines “epistemic cognition,” as cognition about the limits of knowing and the certainty of knowing. Epistemic cognition is a step removed from metacognition; it is cognition about metacognition.

To operate on higher levels of metacognition, it is arguable that a being must be able both to differentiate between ‘self’ and that which is ‘not self’ and to have a conscious concept of ‘mind.’ Premack and Woodruff (1978) report that chimpanzees are capable of choosing, after watching a problem portrayed on
videotape, the picture that represents the solution to that problem from a set of pictures. They conclude that because a chimpanzee can “impute mental states to [itself] and others” (p. 515) in that way, it has a theory of mind. In a similar experiment, Povinelli, Nelson, and Boysen (1989) had a chimpanzee choose the location of a ball after receiving hints from two humans: one who always provided correct answers about where the ball was, and one who guessed its location. The chimpanzee most often took the advice of the human who was always right. The authors attribute this behavior to “the ability to understand that a different visual perspective may lead another individual to possess a different state of knowledge than yourself” (p. 4). Only by distinguishing between ‘self’ and ‘not self’ can this understanding develop. Such findings in work with chimpanzees are reflected in Gallup’s (1983) claim that self-awareness “subsumes both consciousness and mind ... [a self-aware being is] capable of becoming the object of its own attention” (p. 474).

In defining metacognition in terms of the level of consciousness needed, confusion from its expansive index of definitions is avoided and implicit assumptions about what subjects are capable of it are made explicit. By definition, our operational definition of metacognition avoids questions of the necessity of consciousness for metacognition. The metacognition tested in this study was first defined by Smith, et al. (in preparation). They “do not claim to have demonstrated that [metacognitive subjects] have a more explicit consciousness of the self as an entity, in which they are aware of themselves as actors or agents in the task” (p. 11). We make the same claim about metacognition: subjects who behave metacognitively in this study will have successfully monitored their own internal
cognitive state, but that is not to say that they have a conscious awareness of that subjective state or of themselves as individuals. The metacognition this study shows can be classified as a metacognitive experience; the task does not demand the self-reflection and self-consciousness that metacognitive knowledge, metacognitive control, and epistemic cognition entail.

**Smith, Schull, and Washburn’s Task**

A more elaborate description of the task from which the present one was adapted will serve many functions: It will clarify how our task operationally defines metacognition so that the amount of consciousness required is none, and it will make clear how metacognition can be assessed so that no verbal abilities are required. The psychophysical background demands for this work will also be explored.

Smith, et al. (in preparation) study metacognition in rhesus monkeys by providing an escape response to a visual density discrimination which gets progressively more difficult over trials. The contingency is such that a correct response to the primary discrimination is rewarded, an incorrect response punished, and the escape response is rewarded after a brief delay. On easy trials, it is most adaptive to make the primary discrimination and respond accordingly. On difficult trials, when there is a high chance of guessing wrong, it is most adaptive to pay the price of the delay in order to get a guaranteed reward. By making it adaptive to use the escape response only on the most difficult trials, the authors theorize that the animals “use this ‘bailout’ escape response sparingly when they know they are unsure of the correct response, provided that they have the requisite metacognitive abilities” (p. 3).
Psychophysical thinking provides a framework for thinking about this task, but stops short of covering more subjective aspects of decision-making like responding to subjective uncertainty by using the escape response. In the primary density discrimination, subjects in this experiment are forced to make comparisons of equality between a mental image of a certain density and the image on the screen. Fernberger (1929) suggests that detecting equality might be seen as a failure to detect a real difference, and that results can be more dependent on subjective internal stimuli than on the stimuli presented in the experiment: “We now have well standardized conditions covering the more objective phases of most psychophysical work. What we need is similar standardization of the more difficult and more illusive subjective phases” (pp. 111-112). The present study answers Fernberger’s call for a standardization of the subjective phases of perception; uncertainty monitoring and a subjective decision are at the heart of Smith et al.’s (in preparation) task.

Scholnick and Wing (1988) argue that the subjective uncertainty monitoring skills required of subjects in this study might develop later than other skills regarding decidability because “...it is easier to detect when an answer is decidable than to detect when the answer is undecidable” (p. 195) To recognize undecidability, one must have confidence in one’s own lack of confidence that a question is decidable; cognitively and developmentally, more skill is involved in undecidability: “Pieraut-Le Bonniec (1980) noted that dealing with uncertainty involves more than a recognition that there is no basis for making a choice. It also involves recognition of the legitimacy of undecidability” (p. 195). Scholnick and Wing consider this recognition a developmental milestone; subjects may fail to
show metacognition simply because they have not developed the awareness that undecidability is reasonable.

Smith, et al.’s (in preparation) task uniquely combines the demands of the psychophysical literature to explore the subjective aspects of decision-making with metacognition defined without consciousness or language as prerequisites. While language and consciousness of self may be precursors to higher, more self-reflective and linguistic forms of metacognition, the contingencies in this study are set up so that a simple metacognitive experience, not based on conscious reflection or language, takes place in metacognitive subjects.

Language

Having conceptualized metacognition as a broad term denoting a hierarchy of abilities, and having seen where and how Smith, et al.’s (in preparation) operationalization of metacognition fits into that hierarchy, I will now review and challenge assumptions about what linguistic skills are necessary for metacognition. Often, thinkers assume that language severely restricts thought. But like the ability to refer to a conscious self, it is a definitional issue; the ability to use language is a prerequisite for metacognition only if metacognition is defined that way.

Language unquestionably imposes some structure on thought. The argument revolves around how much that structure determines thought. Vygotsky (1962) proposes that “Words play a central part not only in the development of thought but in the historical growth of consciousness as a whole” (153). Whorf offers an even stronger linguistic determinist position; he argues that because the Hopi Indians had no terms for concepts of time, they necessarily had no thoughts
about time (in Schultz, 1990, p. 27). Both Vygotsky and Whorf would argue that language restricts and shapes all cognition, including metacognitive thought.

Alternatively, one can argue that while language is the primary vehicle through which thoughts are expressed, and that language might influence thought, that thought is possible without language. Schultz defines this notion as “linguistic relativity,” explaining that the “...structure of a human being’s language influences [but does not determine] the manner in which he understands reality and behaves with respect to it” (p. 14). Focusing on metacognition, Markman (1981) agrees, holding that: “...it is not always necessary to have an explicit question ‘Do I understand?’ in order to obtain further information about one’s own understanding” (p. 75). Gallup (1983) also believes that language has little control over thought. He requires only that an organism be self-aware and capable of being the object of its own attention to be considered metacognitive; he states very directly, “...language is not a precursor to mind” (1983, p. 506).

In the study of human metacognition, implicit assumptions have been made about how language influences thought. The verbal stimuli and verbal assessments used to study the existence and the development of metacognition in children reflect the unspoken idea that language skill is not a confounding factor in measuring metacognition. But there are a number of problems with that reasoning:

First, it is possible that previous work has ignored certain types of metacognition because it is not at a high enough cognitive level to be measured by verbal means. A metacognitive experience, for example, requires no language, so a verbal assessment will ignore the metacognition if language skills needed to express oneself verbally develop after the ability to have a metacognitive experience. It is
unreasonable to assume that even a nonverbal infant or very young child is incapable of metacognition.

On the other hand, verbal tests of metacognition may overestimate the ability. If verbalization of metacognition is a fundamentally different kind of cognitive activity in that an idea must be abstracted and verbalized, verbal assessments of metacognition might encourage metacognitive thinking by prompting a child who might not otherwise to think in an abstract and metacognitive way.

Verbal reports of metacognition are simply unreliable. Piaget reports that when children speak, they distort and modify both mental and external observations. Work on eyewitness testimony, in which drastically different reports of the same event are obtained from different people, also demonstrates the unreliability of verbal data (in Brown et al., 1982).

The present research uses Smith et al.’s task for the purpose of studying metacognition in children because it is not verbal: “A non-verbal procedure ... could reveal uncertainty monitoring at a very young age if that capacity is present” (Smith et al., in preparation, p. 10). It is clear from the Smith et al. work showing metacognition in chimpanzees that language is not essential to metacognition; the two need to be teased apart in order to fully grasp their interdependence. Building on Smith, et al.'s (in preparation) nonverbal paradigm for testing metacognition will allow us to overcome underestimations and overestimations of metacognitive ability that arise in a verbal experiment.
Metacognition in Humans

With the exception of Smith, et al.'s (in preparation) suggestion that children's metacognition can be objectively tested, previous research has relied almost entirely on verbal means of evaluation. I will review the human metacognition literature from this perspective, showing how standard procedures for testing metacognition have been inadequate. I will first define the two major areas of research, cognition about comprehension and cognition about memory. Then, I will review studies in terms of variables that affect metacognitive behavior such as experience, strategy, structure, and instructions. Finally, I will point to developmental issues and to the hope offered by nonverbal means of assessing metacognition.

Metacomprehension, also called comprehension monitoring, is a conscious reflection on one's own understanding. It holds a prominent place in the traditional metacognition literature because it is applicable to many tasks and it is easily tested. Its usefulness is clear; one who monitors his own comprehension processes will be able to evaluate and modify those processes to increase his comprehension more easily than will one who does not engage in such self-regulation: “individual and developmental differences in sensitivity to one’s comprehension failure will result in differences in the quality of comprehension itself” (Markman, 1977, p. 992).

Metamemory is the other main area of human metacognition that has received much attention in the literature. Kreutzer, Leonard, and Flavell (1975) define metamemory as a “child’s verbalizable knowledge of how certain classes of variables act and interact with one another and affect the quality of an individual’s performance on a retrieval problem” (abstract).
Variables Influencing Metacognitive Behavior.

Flavell and Wellman (1977) suggest that person, task, and strategy variables are critical in determining how likely it is that a child will demonstrate metacognitive behavior. This section reviews those and other such variables to demonstrate the relative advantages of nonverbal inquiry.

One important determinant of metacognitive behavior is experience. Experience breeds expectations, and thus prepares individuals to more carefully monitor their comprehension to see if their expectations are met. Markman (1981) concurs, adding that experience, gained from observing oneself or others, improves comprehension monitoring and may aid in the development of metacomprehension. She proposes that children are generally less successful at comprehension monitoring than are adults not because they are not yet developed, but because they lack experience. She reports that child chess players are better at remembering formations of chess games than are adult novices, claiming that novices are poor monitors of their own comprehension and that children are simply universal novices (p. 76).

Children must also gain experience in terms of developing a concept of 'self,' an understanding of one's own memory limitations and idiosyncrasies, and an ability to read one's own memory states. Markman (1981) found that five-year-olds could predict their performance on motor tasks before they could predict their memory abilities. Kreutzer et al. also points to self-awareness as a function of age: "The older children seemed to have a more differentiated self-concept ... than the younger ones" (in Flavell & Wellman, 1977, p. 12).
The possession of a clear strategy for behaving metacognitively is another determining variable; given a strategy to use, young children who wouldn't otherwise behave in a way suggestive of metamemory, use that strategy successfully and understand why it works. Flavell (1978) proposes an explanation: "...belated reflections could influence future memory behavior, of course, and may in fact comprise one of the processes by which metamemory develops in children." This proposition is based on the knowledge that "young children are likelier to use a memory strategy spontaneously once they have become aware of its usefulness" (p. 230). This explanation focuses on experience and strategy as major catalysts for the development of metamemory abilities. In an experiment dealing with this issue, Wellman et al. asked three-year-olds to remember under which of a number of cups a toy was hidden. Subjects who were told to wait with the toy remembered better than those who were simply told to remember where the toy was. The ready availability of physical, spatial, external strategies for remembering, touching the cup, looking at the cup, and standing next to the cup, enabled very young children to use strategies for remembering (in Flavell & Wellman, 1977, pp. 8-9). Young children may develop metamemory through experience, retrospectively examining their performance.

Strategy variables themselves also need to develop. "Strategies can be beneficial only to the extent that learners anticipate their need, select from among them, oversee their operation, and understand their significance" (Campione, et al., 1982, p. 434). The ability to use strategies develops with age: When Kreutzer et al. (1975) asked children to imagine that they would have to take their skates with them the next day, older children thought of more ways to remember, and their
strategies included greater planning than did those of younger children. All children used external more than mental reminders. Belmont and Butterfield (in Brown, et al., 1982) found that children were capable of modifying their strategies. They studied developmental differences in rote memory and found that older children abandoned strategies when they were no longer necessary and modified strategies when appropriate.

Experimental manipulation also provides variables which alter the likelihood of exhibiting metacognition. Markman (1981) reports that the structure in the material to be evaluated is one such variable. With no structure in which sentences in a paragraph relate to each other, for example, children evaluate for truth at a sentence level, failing to compare consistency between sentences. Other times, a multiple structure is present; two or more interpretations are possible. In such cases, people truly monitoring their comprehension will recognize the ambiguity. Dickson (in Markman, 1981, pp. 67-72) found that children faced with a multiple structure do not realize the ambiguity. He asked children to choose an object that fit an ambiguous description. The children selected the first object they saw that fit the description, as opposed to noting the ambiguity and the alternative correct answers. When a task contains either no structure or a multiple structure, it becomes more difficult and more important to adequately monitor one’s own comprehension.

In discussing task variables, Flavell and Wellman (1977) point out that tasks vary in difficulty: some data are difficult to store, based on familiarity and meaning, some requests are difficult to fulfill, and some variables are irrelevant to successful completion of the task (p. 15). Experimental work has determined a basic pattern for the development of the use of task variables: Yussen and Bird
(1979) show that three-year-olds know that recalling a small set of pictures is easier than recalling a large set (in Flavell and Wellman, 1977). Kreutzer, et al. (1975) show that kindergartners know that increasing the number of items to be memorized increases the difficulty of recalling them. Some six-year-olds and all seven-year-olds knew that more study time increased memory. By age nine children realized that opposites are easier to remember than are unrelated words, and that it is harder to recall names after learning another, different list (pp. 16-19). Kreutzer, et al. (1975) found that while kindergarten and first-grade students used external mnemonic resources and Brown’s person, task, and strategy variables, third- and fifth-grade students were “considerably more planful and self-aware in their approach” (abstract). They had a more differentiated concept of self and they better understood relations among items to be remembered. Wellman (in Brown, et al., 1982, p. 108) reports that an understanding of the relevance of variables occurs in this order: number of items, distraction, age of rememberer, assistance from others, study time, and associative cues.

Instructions also constitute a variable that affects the success of comprehension monitoring. This research is very practical; specific instruction in metacognition has been found to improve students’ performance (Paris & Oka, 1986; Markman 1979). Markman (1979) found that third-graders she studied seemed to be ignoring obvious inconsistencies in essays they read. While they may have been capable of comprehension monitoring, their success at finding and reporting inconsistencies was affected by the instructions they received: those who were told to listen for obvious inconsistencies in the essays were more likely to report noticing them than were controls who were not told beforehand what their
task would be. Markman hypothesizes that the number of simultaneous processes is prohibitive, and that a failure of metacomprehension may simply reflect the need to develop the necessary skills: "children have to encode and store the information, draw the relevant inferences, retrieve and maintain the (inferred) propositions in working memory, and compare them" (p. 643). She maintains that children may not be able to accomplish all of those skills without explicit direction even if they are capable when given direct instructions.

Flavell and Wellman (1977) make a similar argument regarding developing skills to perform metamemory tasks; their "differentiation hypothesis" proposes that the developing child "...must learn to differentiate a future-oriented memorization instruction from a present-oriented perception instruction" (p. 7). Appel studied four-, seven-, and eleven-year-old's conceptual and behavioral differentiation, and found that while the four-year-olds were not differentiating, the seven-year-olds were (in Flavell & Wellman, 1977, pp. 7-8).

Cognitive development strongly influences the ability to behave metacognitively. The basis for metamemory is detectable in children's behavior before metamemory is: "...preschoolers possess basic monitoring abilities that are important for the development of subsequent skills such as the allocation of mnemonic effort" (Cultice, Somerville, & Wellman, 1983, p. 1480). In judgements of both feeling-of-knowing and of novelty of a stimulus, children "...predict subsequent recognition performance, indicating accurate memory monitoring" (p. 1480) before behaviors indicative of metamemory are entirely developed and demonstrated.
Flavell (in Brown et al., 1982, p. 115) had children construct block buildings based on ambiguous directions given by a confederate child on a tape. Children were allowed to rewind the tape to listen again, and were asked later to evaluate the directions. Younger children did not rewind the tape, show confusion, or report the ambiguity in the directions. Markman (1977) theorizes that an absence of constructive processing, an ability to separately and consciously monitor one’s own comprehension in an “active, self-directive” manner (p. 986), is responsible for children’s poor performance on tasks like Flavell’s, and hypothesizes that it is fundamental to successful comprehension monitoring: “...children appear to be processing the material at a relatively superficial level, not really attempting to execute the instructions mentally or determine the relationship between the instructions and goal” (p. 991). Constructive processing is yet another variable that needs to develop.

There are many developmental explanations for a failure to observe behavior indicative of metamemory or metacomprehension. None imply a lack of ability. First, one may not use metamemory despite having the capability to; verbal questioning might elicit responses indicative of the capability of metamemory. Second, metamemory may not be well enough developed to use, in which case questioning will not elicit such responses. Third, metamemory may occur reflectively, in which case it comes too late to be used to govern behavior (Flavell, 1978, p. 229).

Wellman (1977) presents a study which illustrates this phenomenon. He found that young children who knew whether or not they had seen a pictured face before were poor at predicting how many faces they would recognize. He outlines
a few potential explanations for why younger children don’t predict their abilities accurately (p. 20-21), each attributing more metacognitive ability: First, it is possible that the children didn’t understand “what the feeling of knowing judgement required” (p. 19); they didn’t understand the directions, or they were inexperienced in making feeling of knowing judgements. Second, their memory system may be organized differently, so there is no relevant information for them to access. Third, the relevant information may be there, but they are not capable of accessing it. Fourth, they are capable of accessing the information but they simply don’t think to do so. Wellman hypothesizes that this explanation is the most likely to be correct, because questions were asked separately and were not connected in the children’s minds. Finally, it is possible that the information exists, children are capable of accessing it, they do, and they simply do not differentiate between useful and useless information. Clearly, the failure to observe behavior governed by metacognition does not imply that the individual is incapable of metacognition.

Nonverbal tasks.

More recent studies have used less verbal and more ecologically sound techniques, avoiding difficulties with verbal data. Revelle, Wellman, and Karabenick (1985) had experimenters play with three-and four-year-olds and make both easy and difficult requests. They report that “...young children [four-year olds, but not three-year-olds] are capable of detecting a variety of comprehension problems and possess strategies for resolving these difficulties with a partner in communication” (p. 654). DeLoache and Brown (1984) also used nonverbal means of assessment and were able to find an even lower age for the development of metamemory; they tested 21- and 27-month-olds by hiding a toy and asking the
child to find it. On occasion the child would make an error and forget where it was hidden (an error trial), and on other occasions, the experimenter would move the toy without letting the child know (a surprise trial). At the basic cognitive task, "Both age groups showed greater persistence in their initial search on surprise than on error trials, indicating that their retrieval effort was based on their level of subjective uncertainty" (p. 37). But on surprise trials, they search differently: the older group “searched selectively and intelligently, confining their search primarily to locations that were nearby or in some way related to the place where the toy had [originally] been hidden .... [indicating] that they had made intelligent guesses about plausible alternative locations for the missing toy” (p. 37). Such nonverbal assessment of metacognition sets the background for evaluating performance on the present task.

A Nonverbal Psychophysical task: Summary and Study Questions

This study will tie together many questions that have been raised in different areas in the literature. If we can show that a low-level metacognition exists despite a lack of verbal ability, we can legitimately question the literature’s implicit assumption that language is a precursor to all metacognition. This work will also serve as a complement to the comparative data pointing to the similarities between that the human mind and the animal ‘mind.’ In applying Smith et al.’s specific definition of metacognition to the human literature, we have introduced a new, more systematic and objective test for metacognition.

Our main hypothesis is that verbalizing about metacognition is more developmentally advanced than simply behaving metacognitively. Verbalizing requires more conscious reflection as does metacognitive knowledge. By ignoring
that fundamental difference, most traditional metacognitive literature has ignored an ability to behave metacognitively just because it is not paired with an ability to think consciously and to speak about metacognition.

To test this hypothesis, we will first examine the discrimination and metacognition behaviors and verbalizations to determine whether or not the ability to speak about metacognition operates at a different level. Second, we will examine developmental level as reflected by age and class to determine when the ability to speak about metacognition develops relative to the behavior and to the more basic cognitive tasks of discriminating and speaking about one’s discrimination. We hypothesize that verbalization of metacognition is a higher-order ability that develops later and operates on a higher level of consciousness. We believe that verbal measures will underestimate the presence of metacognition, implying that metacognition may exist in younger children than the traditional literature suggests. We will test these hypotheses using an adaptation of Smith, et al.’s (in preparation) task, a task more likely than the traditional literature to elicit behavior demonstrative of metacognition if metacognition exists due to its specific reward and time-out contingencies.

Method

Subjects

The children in the experiment attended a private nursery school in an affluent suburb of Philadelphia. Their parents gave permission for their participation in advance of the study. Subjects were told they were going to play a
computer game and that they would have a chance to receive stickers depicting animals and cartoons.

Students in the nursery school's two classes formed the subject population for this study. A total of 28 subjects participated. Of the 21 who provided fair data\(^2\), 12 students were in the older class, with an average age of 5.0 (standard deviation = 0.2), and 9 students were in the younger class, with an average age of 4.2 (standard deviation = 0.3).

Subjects were selected for participation by the experimenters in consultation with the nursery school teachers based on these factors: the original intention to include all the children, the likelihood that they would be most attentive to the task, and the subject's desire to participate.

**Apparatus**

We used an IBM compatible computer and a monitor, an EloGraphics touch-screen that operates by sensing the interruption sound waves, a standard monitor, small pleasing stickers of animals and cartoons, a video camera, and a kitchen timer.

**Procedure**

During the weeks preceding the experiment, we spent approximately eight hours as observers in the subjects' classroom to encourage the children to feel comfortable in our presence. In the older class, a program in which one 'paints' on the screen using a finger was made available to everyone in the class during the 'activities period,' an unstructured part of the day when children were allowed to

---

\(^2\) A Subject's data was excluded if he or she was never offered the escape response, if he or she participated as part of a pair before we asked questions, or if he or she was never asked questions because of other constraints.
choose from a variety of activities. Only some of the subjects used the computer at that time.

For the experiment, we took individual children from their classrooms to another room and told them that they would be able to win stickers by using the computer touch-screen to earn slips of construction paper. Ten slips of paper earned a sticker. Each individual participated in between one and five sessions. Questions not related to strategies for doing well in the game were answered throughout the session.

The computer program displayed at most three objects on the screen: a variable number of pixels randomly spread out across a 100 x 200 pixel box, an ‘S’, and a star. The stimulus for determining the correct response was the number of pixels in the box. With 2950 pixels, the trial was considered dense, and touching the box was the correct response. With any fewer pixels (450 to 2949), the trial was considered sparse, and touching the ‘S’ was correct. Touching the other primary stimulus was incorrect\(^3\).

On every correct response throughout the experiment, the computer emitted a rewarding sound, showed overlapping rectangles of different colors, added a square to a row of squares on the bottom of the screen, and started a new trial immediately. This reward sequence was complemented by the experimenter’s verbal encouragement with reinforcing words such as “good job!” or “very good!” After a series of correct responses, a row of squares on the bottom reached across the width of the screen, colorful randomly placed lights were shown on the screen, and

\(^3\)At one point, we used a standard monitor and simulated the effects of the touch-screen by allowing the subjects to touch the screen while the experimenter entered their responses into the computer using the keyboard.
and a slip of construction paper was given to the subject. As a punishment on incorrect responses, a buzzing noise was emitted and there was a delay in the onset of the next trial.

The program shaped the subjects to learn the density discrimination by increasing the difficulty of the task. Only after each stage was mastered, defined as responding correctly five times in a row, did the program proceed to the next. It started by presenting only one stimulus at a time (first the dense box and then the ‘S’) where the correct response is to touch that stimulus. Then, it presented two stimuli: First, the most sparse box and the ‘S’, and then the dense box and the ‘S’. Finally, it presented the ‘S’ and either the most sparse or the dense box. Subjects who were having trouble acquiring the discrimination were given verbal suggestions such as “It has something to do with the number of dots.”

Once the primary discrimination was mastered, the program displayed sparse trials that, from one trial to the next, become more and more dense, tracked the subject’s success, and adjusted the next trials: making the discrimination more difficult if the subject is responding correctly more than 50% of the time and making it easier if the subject is responding correctly less than 50% of the time.

As the sparse trials became more dense, a star appeared on the screen. When the star was touched, the next trial had only one primary stimulus on the screen, and a win was guaranteed. But, in order to discourage constant escape-behavior, repeated use of the star bore a cost: the guaranteed-win trial was delayed more and more as the star is overused. To make that contingency clear, a circle, centered at the center of the star, was added to the screen. Its size decreased with every primary discrimination response. The circle served as a visual and auditory
display that there was a delay before the guaranteed win was make available: during that delay the circle got bigger and then shrank slowly, accompanied by a single repeating tone to indicate the delay. Moreover, a timer was set at the beginning of the session to show children visually that the delays due to overusing the star are in fact costly.

In order to make the function of the escape response clear to subjects, those who were not using the star at all were presented with trials in which the program would only proceed after the subject uses the star. These trials were not included in the analysis of spontaneous escaping, and they were the only encouragement to use the star; verbal assistance was given about the primary discrimination, but the experimenters did not give verbal cues about the escape behavior.

At the end of the final session, or when the experimenter thought that the child was behaving metacognitively, each subject was asked to verbalize his or her strategy for winning stickers. Specifically, they were asked when they touched each of the three objects on the screen. If a clear answer was not given, the question was rephrased repeatedly until the experimenter was satisfied that the subject would not respond in a way clearly indicative of an ability to verbalize strategies for touching the three objects. All interviews were recorded on videotape or audiotape.

Some subjects participated in pairs in order to encourage verbalization about their metacognition. In these special cases, the two needed to agree on what object to press in order to win stickers. The same contingencies were provided, and each subject won a sticker whenever the other did. While these interviews were recorded for evidence of more naturalistic verbalizations, none of the data collected
from pairs or from individuals who had previously been part of a pair was included in the data analysis.

Results

Behavioral scores

Each subject's data were scored along two behavioral dimensions: ability to perform the discrimination behavior (touching the box on the dense trials and the S on the sparse trials) and ability to perform the metacognition behavior (touching the star on the most difficult discrimination trials).

For the discrimination behavior, subjects were assigned a discrimination index determined by combining the percent of correct responses on dense trials and the percent of correct responses on sparse trials in a ratio of 3 to 1. Since pure dense trials are more difficult than pure sparse trials, they are more telling of discrimination ability, and were therefore given more weight. The average of each subject’s percent correct on dense trials was 68.5 (standard deviation = 17.1), and on sparse trials was 92.3 (standard deviation = 13.3). Figure 1 is an example of one subject’s discrimination performance.

Insert Figure 1 about here

On sparse trials (trial difficulty = 21), the subject responded correctly by pressing the ‘S’ 100% of the time. On dense trials, the subject responded correctly by choosing the box 78% of the time. Based on this index, subjects were ranked in order of their discrimination ability.
Subjects were also ranked for metacognition behavior. Three experimenters individually scored the behavior by reviewing the data. They were reliable; the lowest correlation between any of the three pairs of reviewers was 0.80. The experimenters’ individual rank-order score for each subject were added together to compute a final non-parametric rank-order for behavioral metacognition. Figures 2, 3 and 4 provide examples of graphs for progressively better metacognition. Figure 2 shows almost no starring behavior. Figure 3 shows more starring behavior at the most difficult trials, but the behavior is not centered at one difficult discrimination level. Figure 4 reflects the most metacognitive ability; the star is used exactly when the subject is most confused and equally likely to be correct or incorrect at the discrimination.

Verbalization scores

In addition to the behavioral scores, the verbalizations of each subject were reviewed for utterances indicative of the possession of the ability to discriminate or the ability to behave metacognitively.

Three reviewers independently examined the transcripts of the sessions for any instance of verbalizations of strategies. In regard to verbalizing an underlying strategy for responding correctly to the primary discrimination, subjects were grouped into two categories: those who did (score of 1) and those who didn’t (score of 0). Subjects were also scored for verbalizing a strategy for when they touched the star; they were grouped into three categories. Those who clearly used
metacognitive language were scored 2, those whose verbalizations were ambiguous with regard to whether or not they were metacognitive were scored 1, and those who did not use metacognitive language were scored 0. There was significant reliability between the reviewers: the lowest correlation between any of the three pairs of reviewers was 0.73 for verbalizations of the discrimination and 0.60 for verbalizations of the metacognition. Scores were then reviewed collectively until agreement could be reached between the reviewers. Typical verbalizations for each non-zero score level are included in the Appendix.

All data, including behavioral rank orders for the discrimination and the metacognition, the verbalization scores for the discrimination and the metacognition, class, age, experience as number of sessions, and sex is included in Table 1.

Correlations

One-way pairwise Spearman rank order correlations were computed to analyze the above rankings with the independent measures of sex, experience, developmental age (measured by class) as well as with each other. Table 2 contains a correlation matrix and a matrix identifying the number of subjects in each cell.
In order to determine where metacognition and the ability to verbalize about metacognition fit into the hierarchy of metacognitive abilities, and to see where they stand developmentally, we examined the behaviors and the evidence of their verbalization as they are predicted by age and by each other. Students in the older class were more likely to have good behavioral discriminations ($r = 0.46, N = 21 \ p < .025$) and, to a marginally significant level, to be able to explain that discrimination verbally ($r = 0.41, N = 17, p < .1$). While age did not predict demonstration of the metacognitive ability ($r = 0.10, N = 21, p > .05$), it did strongly predict the ability to verbalize about metacognition ($r = 0.68, N = 17, p < .005$). Ability to perform behaviorally on either the discrimination or the metacognition did predict the other ability ($r = 0.41, N = 21, p < .05$), but the ability to verbalize about one did not predict the ability to verbalize about the other ($r = 0.30, N = 17, p > .05$).

We tested the accuracy of verbal reports by examining each behavior with its corresponding verbalization. The ability to discriminate predicted the ability to verbalize about the discrimination ($r = 0.47, N = 17, p < .05$), but the ability to behave metacognitively did not predict the ability to verbally explain one’s strategy for doing so ($r = 0.35, N = 17, p > .05$).

Sex did not predict any of the abilities measured in this study. The data indicate that girls and boys were generally equal in their behaviors and in verbalizing their strategies in regard to those behaviors. Sex was not significantly correlated with the discrimination behavior ($r = 0.32, N = 21, p > .05$), the metacognition behavior ($r = 0.30, N = 21, p > .05$), the ability to verbalize strategies for the discrimination behavior ($r = 0.38, N = 17, p > .05$) or the ability
to verbalize strategies about the metacognition behavior ($r = 0.39, N = 17, p > .05$). There was not a significant difference between the two classes in terms of sex ratio ($r = 0.36, N = 21, p > .05$).

Experience with the experiment, measured by how many sessions a subject participated in, was not significantly correlated with class ($r = 0.13 \ N = 21, \ p > .05$), the behavioral discrimination ($r = 0.10, \ N = 21, \ p > .05$), the behavioral metacognition ($r = 0.30, \ N = 21, \ p > .05$), the ability to verbalize strategy about the discrimination ($r = 0.38, \ N = 17, \ p > .05$), or the ability to verbal strategy about the metacognition ($r = 0.31, \ N = 17, \ p > .05$).

**Discussion**

**Major contributions**

We have verified that the primary discrimination task is reflective, both behaviorally and verbally, of developmental stage. The behavior and the verbalization are marginally correlated with class, suggesting that they are both measures of the developmental progression we have assumed to be approximated by class in school. The ability to behave metacognitively is just as likely to be found in members of both classes, suggesting that it may develop in even younger children. With a nonverbal paradigm in place, future research is now able to look at metacognition in even prelinguistic children.

The ability to talk about one's metacognitive behavior, however, tells a different story. Children in the younger class, who are just as likely as children in the older class to behave metacognitively, are not nearly as likely to speak about their metacognition; not one subject from the younger class verbalized about his or her metacognition. This finding suggests that the ability to verbalize about
Metacognitive Speech in Children

metacognition develops later than the ability to show behavior indicative of metacognition. Further, the two behavioral measures, one requiring cognitive abilities and the other requiring low-level metacognitive functioning, are correlated, while the abilities to verbalize about those behaviors, one about a cognitive ability and one about a metacognitive ability are not correlated. These findings suggest that the lower-order metacognitive experience that we've demonstrated is indeed lower than the ability to explain one's behavior in a hierarchy of increasing consciousness and cognitive ability. Verbalization requires a more conscious, self-reflective analysis of one's behavior than does simply behaving adaptively; it represents a step up on the hierarchy from metacognitive experience to metacognitive knowledge. By ignoring that fundamental difference, most traditional metacognitive literature has ignored an ability to behave metacognitively just because it is not paired with an ability to think consciously and to speak about metacognition.

Another major finding relates to the validity of verbal self-report data in determining the level of metacognitive functioning in children. It was hypothesized that verbal data, the most frequent means of measuring metacognition, is invalid. The fact that metacognitive behavior and metacognitive language were not correlated strongly suggests that different abilities are being used in speaking about one's own internal states and acting on them. Some children are able to perform metacognitively on this task but are not capable of verbally expressing their strategies for their adaptive behavior. Using this nonverbal technique to measure metacognition, it was possible to separate the behavioral skill from the verbal skill, and to call into question the accuracy of verbal measures.
Sex and experience, potential confounding factors, were not significantly correlated with any of the dependent measures, eliminating the possibility that they confounded the results.

**Experimental Inconsistencies**

Several potential sources of noise were related to the subjects. To begin with, we had only seventeen subjects with complete data and we were unable to randomize our subject population. Because some subjects refused to participate, others did not do well, teachers would only let certain children leave the classroom at a given time, and there were time limitations, our selection of subjects was constrained by both random and nonrandom factors. Another source of disparity may have come from subjects' discomfort, either from having never used a computer before, or from being asked to play a game with relatively unfamiliar people watching.

Many potential sources of noise were due to the fact that the experimenters attempted to give each subject the best chance to perform metacognitively. Lack of consistency was often intended to help each subject learn the discrimination and to provide the optimum chance that any metacognitive ability would be demonstrated behaviorally. The experimental technique was highly variable and a potential source of error because it developed over the course of the experiment. The demeanor of the experimenters became more verbally interactive and was, in general, variable from one experimenter to another in terms of words of encouragement, frequency of encouragement, interactive style, and wording in asking follow up questions to draw out those who did not respond to the first. The matching of experimenter to subject was not randomized.
The experiment itself was also flawed. Inconsistent environmental situations were created because different stickers were available as rewards, and neither the video camera nor the timer was always used. The number of correct responses needed to get a piece of construction paper (one-tenth of a sticker) was different depending on how interested the subject was in continuing. When it became necessary to simulate a touch screen by observing the subjects and entering responses into the computer with a keyboard, some small amount of subjectivity came into play.

We set up contingencies to define the metacognitive behavior as adaptive, but some evidence suggests that our rewards and time delays were not doing what we intended. We assumed that winning stickers was the only motivation for the subjects. On the contrary, the children were often rewarded just by having the opportunity to play with a computer, to talk to the experimenters, and to focus on their distracting surroundings. In fact, even the time-outs, which need to be negative reinforcement for the experiment to work, were sometimes appealing. For example, subjects sometimes enjoyed watching the circle which was meant to visually display the time delay.

**Lower-level Explanations**

Smith, et al. (in preparation) were able to do more fine tuning of their experimental manipulation than we were. With the goal of refuting criticism that their findings could be dismissed by lower-level stimulus-response explanations, they measure time delay in a subject's responding and watch how subjects respond to an increased cost of escaping the discrimination trial. There are basically three lower level interpretations of the data; each says that the children are simply
responding to an objective stimulus, and each needs to be addressed (Smith et al., in preparation).

First, the argument can be made that, on difficult trials, the cognitive indecision causes hesitation in the form of a long latency period between the onset of the trial and the time a decision is made or some other kind of objective stimulus about the cognitive behavior itself. This latency might in itself be an objective signal to the child that the star is the best response. To test this notion, one would examine trials during which there were long latency periods. Smith, et al. (in preparation) found that monkeys do not select the star significantly more often than either of the primary stimuli after long latency periods, and that on those trials where the star is chosen, there is not a significantly longer latency period than on trials when one of the primary stimuli is used. To test this idea in children, one would first have to find a way to distinguish latency due to indecision from latency due to distraction, a likely confounding factor in human subjects who are not depending on success in the experiment for food.

Second, one might say that the psychophysical stimulus range has simply been divided into three ranges and that the star is considered the best response to the objective middle stimuli. But, this argument is countered by the reward and punishment contingencies set up in this experiment: if it were truly the case that there are three ranges, each with its own response, the animals would choose the ‘S’ when presented with a density in that supposed middle stimulus range because that response would bring a reward with no waiting period. Therefore, it can not be the case that there are three objective stimuli ranges.
The third argument might hold that there are a range of subjective impressions created by the objective stimuli, and that, for the middle range, the star is the best response, but that metacognition is not present. But if that were the case, as the punishment for using the star increased, one would expect the use of the star to decrease overall, and to be more concentrated at the same level of trial difficulty. This variable was not manipulated in this study, but Smith, et al. (in preparation) found that subjects reserved the star for the more difficult trials, not for the same level of trial difficulty, showing that the star use is necessarily due to metacognitive functioning.

**Future Research**

One way to expand upon the findings in this study is to repeat Smith, et al.'s (in preparation) procedure with children. By manipulating the cost of using the star and examining the trials during which there are long latencies, one would be able to verify Smith, et al.'s (in preparation) primate results with children.

Considering that this experiment uses a distinctly different approach in measuring metacognition, insofar as it is nonverbal and independent of verbal report, subjects are not evaluated in a way comparable to the traditional metacognition literature. To have a comparable perspective, future research should also administer a more mainstream task. If the age at which metacognition develops is found to be different, we would have a stronger argument to support the existence of a metacognitive hierarchy and more reason to question the use of verbal data as a measure of metacognition.
Implications

This research has implications for a wide variety of questions. We have shown that metacognition, more specifically, a metacognitive experience, can take place in children at an age where the traditional literature has failed to show metacognition. Clearly, metacognition is not as directly linked to language as it might appear from more traditional work. It follows that, in using verbalizations or a lack of verbalizations about metacognition to draw conclusions about all metacognitive functioning, psychology has missed part of the story.

This work also serves as a complement to comparative research. By comparing the results of this experiment with similar results in other species, the phylogenetic origins of metacognition as an adaptation can be examined. In addition, the theory that human minds are qualitatively different from animal minds can be analyzed. If different species can be shown to perform the same metacognition observed in humans, it follows that their minds are not distinctly different.
Works Cited


Works Consulted

Angell, F. (19??). On Judgements of “Like” in Discrimination Experiments.  

Bondy, E. (1984). Thinking About Thinking: Encouraging Children’s Use of 

Metacognition. In S. R. Yussen (Ed.), *The Growth of Reflection in 
Children* (pp. 105-144). New York: Academic Press.

In B. Gholson & T. L. Rosenthal (Eds.), *Applications of Cognitive 


Problem of Metacognition. In R. Glaser (Ed.), *Advances in Instructional 
Psychology (Vol. I)* (pp. 77-165). New York: John Wiley & Sons.

Development in Educable Retarded Children. In R. V. Kail & J. W. Hagen 
(Eds.), *Perspectives on the Development of Memory and Cognition* (pp. 


Crook, J. H. On Attributing Consciousness to Animals.  

Davidson, D. (19??). *Inquiries into Truth and Interpretation*.  


Human Performance (Vol. I): Theoretical Perspectives. (pp. 57-110).


Metacognitive Speech in Children

Appendix 1

Sample Verbalizations

For the verbalizations about the discrimination, subjects were coded 1 when they did verbalize strategies about their metacognition and 0 when they did not. Some verbalizations that received a score of 1 were:

- "[I touch the rectangle] when it has more dots than the S."
- "[I touch the S] if it has more dots than the rectangle."
- "See when there’re dots on it [referring to the unlit gaps between the lit pixels on a sparse trial], I push the S."
- "[I touch the box] when it has more dots."
- "[I touch the S] when it has less dots."

For the verbalizations about metacognition, subjects were coded 2 when they verbalized strategies about when to choose the star, 1 when such verbalizations were ambiguous, and 0 when they did not verbalize about the metacognition. A verbalization that received a score of 1 was:

- What happens when you touch the star?
  “I don’t know” ...
- Can you tell me why you pressed it?
  “to have one” ...
- Why do you like it [the star]?
  “cause it makes it easier”

Some verbalizations that received a score of 2 were:

- When do you touch the star?
  • “if you’re not sure which one is right”
  • “When you get confused”

- Why did you touch the star that time?
  • “I didn’t know which one is wrong.”
Author Notes

This paper was prepared as a thesis for Haverford College’s Senior Research Tutorial in Biological Psychology. The author worked with two other students, Daniel Karpf and Aaron Vance, and was advised by Professor Jonathan Schull of Haverford College.

We gratefully acknowledge the assistance of David Smith of the State University of New York at Buffalo and Robert Wozniak at Bryn Mawr College for their assistance. We are also deeply indebted to Marilyn Henkelman, Elizabeth Bradley, and Elizabeth Nader of the Phebe Anna Thorne school in Bryn Mawr, Pennsylvania.
## Table 1

### Raw Data

<table>
<thead>
<tr>
<th>Subject</th>
<th>Discrim&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Metacog&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Verbal Disc</th>
<th>Verbal Meta.</th>
<th>Class&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Experience&lt;sup&gt;c&lt;/sup&gt;</th>
<th>Sex&lt;sup&gt;d&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>5</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>20</td>
<td>54</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>17</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>15</td>
<td>31</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>5</td>
<td>47</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>54</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>7</td>
<td>44</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>6</td>
<td>13</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>4</td>
<td>48</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>15</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>12</td>
<td>16</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>13</td>
<td>8</td>
<td>45</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>16</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>19</td>
<td>31</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>35</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>17</td>
<td>18</td>
<td>40</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>8</td>
<td>29</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>19</td>
<td>14</td>
<td>21</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>21</td>
<td>51</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>21</td>
<td>8</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

Note. All empty cells represent excluded data. A subject's verbal data was excluded if he or she was never offered the escape response, if he or she participated as part of a pair before we asked questions, or if he or she was never asked questions because of other constraints.

<sup>a</sup>These are non-parametric rank-orders.

<sup>b</sup>Class is coded 1 for older class, 0 for younger class.

<sup>c</sup>Experience is measured in number of sessions.

<sup>d</sup>Sex is coded 1 for males, 0 for females.
Table 2

Correlation Matrix and Frequency Table

<table>
<thead>
<tr>
<th></th>
<th>Discr.</th>
<th>Meta.</th>
<th>VrbDsc</th>
<th>VrbMet</th>
<th>Class</th>
<th>Exper</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition</td>
<td>0.405</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Discrimination</td>
<td>-0.465</td>
<td>-0.562</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Metacognition</td>
<td>-0.528</td>
<td>-0.348</td>
<td>0.304</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>-0.461</td>
<td>-0.095</td>
<td>0.408</td>
<td>0.675</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>-0.098</td>
<td>-0.299</td>
<td>0.383</td>
<td>0.312</td>
<td>0.131</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>0.318</td>
<td>0.302</td>
<td>-0.383</td>
<td>-0.392</td>
<td>-0.361</td>
<td>0.204</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Discr.</th>
<th>Meta.</th>
<th>VrbDsc</th>
<th>VrbMet</th>
<th>Class</th>
<th>Exper</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metacognition</td>
<td>21</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Discrimination</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Verbal Metacognition</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Class</td>
<td>21</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Experience</td>
<td>21</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>21</td>
<td>21</td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td>21</td>
<td>21</td>
<td>17</td>
<td>17</td>
<td>21</td>
<td>21</td>
<td>21</td>
</tr>
</tbody>
</table>
Figure Captions

**Figure 1.** An Example of the Discrimination Index: Trial Difficulty vs. Response Proportion for one subject.

**Figure 2.** An Example of a Poor Behavioral Metacognition: Trial Difficulty vs. Response Proportion for one subject.

**Figure 3.** An Example of a Moderate Behavioral Metacognition: Trial Difficulty vs. Response Proportion for one subject.

**Figure 4.** An Example of a Good Behavioral Metacognition: Trial Difficulty vs. Response Proportion for one subject.
Figure 1
Figure 2

Trial Difficulty

Response Proportion

Dense 50:50

0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0

21 26 35 44 53 62 71

Dense 50:50
Figure 3
Figure 4