National Library of Canada
Acquisitions and Bibliographic Services Branch
395 Wellington Street
Ottawa, Ontario
K1A 0N4

Bibliothèque nationale du Canada
Direction des acquisitions et des services bibliographiques
395, rue Wellington
Ottawa (Ontario)
K1A 0N4

Your file Votre référence

Our file Notre référence

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

Canada

Gender Categorization and Labeling in 18-Month-old Infants

Alison Derbyshire

A thesis

in

The Department

of

Psychology

Presented in partial fulfilment of the requirements for the degree of Master of Arts at

Concordia University

Montreal, Quebec, Canada

July, 1992

ISBN 0-315-80954-X

Abstract

Gender Categorization and Labeling in 18-Month-Old Infants

Alison Derbyshire

The hypothesis that children begin to organize knowledge about gender only after they are able to apply gender labels to themselves and others accurately, has been called into question by recent research findings. Using the preference looking paradigm, it has been demonstrated that most infants can match voices and faces by gender by the end of their first year of life. In general, comprehension of gender labels has been found to emerge late in the second or early in the third year. The ability to categorize people by gender on the basis of auditory and visual cues, therefore, appears to emerge before infants are able to understand gender labels. In the present study one group of 18-month-old infants were tested for both their ability to match voices and faces by gender and their ability to understand gender labels using the preference looking paradigm. Subjects were presented with a series of trials where they saw pairs of computer digitized male and female faces on two computer screens. In one condition they heard male and female voices calling out to them, and they were required to match the male and female voices with the male and female faces. In the other condition they heard a gender neutral voice asking them to look at the 'lady' or to look at the 'man', and were required to match the gender

labels with the faces. Approximately 65% of the subjects were able to match voices and faces by gender. Among the infants who could match the voices and faces, half of the girls and one-quarter of the boys were able to understand the gender labels. Overall, 47% of the girls and 17.6% of the boys were able to understand the gender labels. The hypothesis that intermodal knowledge about gender precedes comprehension of gender labels was supported. The results of this study suggest that the preference looking paradigm is a more sensitive test of infants' gender knowledge than other methods currently available.

Acknowledgements

I would like to express my special thanks to both Lisa Serbin and Diane Poulin-Dubois for their continuous support, encouragement and input throughout this entire project. I would also like to thank Dr. Dale Stack for her helpful comments. I am indebted to Brenda Kenyon, Ingrid Ostling, and Lucy Brunetti. Without their help, support and encouragement, completion of this project would have been extremely difficult. I would also like to thank Kimberly Powlishta and Jacky Boivin for their help with the statistical analyses. Finally, I would like to thank Paul Smerek who was always there when I needed him, who helped me in every way he could, and who never complained throughout the entire process.

This reasearch was supported by a grant from the Social Sciences and Humanities Research Council of Canada to Drs. Lisa Serbin and Diane Poulin-Dubois. While conducting this research I was supported by a Concordia University Fellowship, a scholarship from the Center for Research in Human Development and by a fellowship from Les Fonds pour la Formation de Chercheurs et l'Aide a la Recherche (FCAR).

Table of Contents

List of Figures

List of Tables

List of Appendices

An extensive literature in the area of sex-role development and sex differentiated behaviour has accumulated over the past twenty-five years. Much is now known about the acquisition of gender related abilities during the preschool and early school-age years. For example, children appear to develop rudimentary ideas about the sex-typing of toys and clothing sometime between 2 and 4 years of age (Weinraub & Brown, 1983; Weinraub, Clemens, Sockloff, Ethridge, Gracely, & Meyers, 1984). Children this age have also begun to form stereotypes about the appropriateness of different activities for members of each sex (Kuhn, Nash & Bruchen, 1978). In addition it has been demonstrated that the well known tendency for children to play in gender-segregated groups also begins to emerge during the third year of life (Maccoby, 1988). A wealth of evidence, therefore, seems to exist indicating that children acquire considerable knowledge about gender at very young ages.

In contemporary research there has been an emphasis on the role of cognitive factors in the acquisition of sex-typed behaviour and sex role stereotypes. The idea that sex-role acquisition is related to knowledge about gender was first proposed by Kohlberg (1966). According to Kohlberg's (1966) model, children acquire knowledge about gender in a series of stages. The first of these stages is gender identity. A child at this stage has the ability to correctly identify him-

or herself and others as male or female. In the second stage, called gender stability, the child understands that he or she has always been the same sex. The final stage of gender constancy is achieved when the child understands that gender is a fixed attribute that is not altered by situational changes.

Kohlberg proposed that the acquisition of sex typed attitudes and behaviours is motivated by a desire to act in accord with one's self perceived identity as a boy or a girl. Self-categorization as a member of one sex or the other has, therefore, been proposed as a pre-requisite for the beginning of sex-typing (Kohlberg 1966; Maccoby 1980). Based on these and similar proposals, it has been argued that sex-typed behaviour and attitudes should develop only after a child has reached the stage of gender identity.

Traditional standardized tests of gender identity are typically passed at approximately 3 years of age (Slaby & Frey, 1975). On these tests children are asked to identify the gender of a series of dolls and photographs and their own gender. For each item the child is first asked "Is this a girl or a boy?" or "Is this a man or a woman?" After the child responds he or she is then immediately challenged with a counter question. For example, if a child originally answered that an item was a boy, he or she would then be asked "Is it a girl?" In order to pass an item a child must correctly answer both the original question and the counter

question, and in order to be recognized as having attained the stage of gender identity he or she must answer all of the items correctly (Slaby & Frey, 1975). The stringent requirements for passing these standardized tests of gender identity may, however, cause us to overlook emergent abilities in infants and toddlers (Leinbach & Fagot 1986). The use of the counter questions may cause many young children to fail, not because they are unable to identify the sexes, but because they are not mature enough to understand the purpose of the counter questions. In their day to day lives, children's answers are usually challenged only when they are wrong. They may, therefore, interpret the counter questions as indications that they have answered incorrectly. This misinterpretation of the meaning of the counter questions may lead many young children, who are able to identify their own gender and the gender of others, to change their correct answers. In fact, according to Leinbach and Fagot (1986) standardized tests of gender identity have "tended to mask early knowledge and to leave children under three looking curiously incompetent with regard to gender" (p.656).

Several studies over the last fifteen years have attempted to simplify the requirements of testing methods, and thereby eliminate as many extraneous performance demands as possible. Using these less demanding methods abilities and knowledge with regards to sex differentiation and labeling have been demonstrated late in the second and early in the

third year. For example, Weinraub et al. (1984) asked 24-to 38-month-olds to sort female and male pictures of adults and children into two boxes, one for men and boys and one for women and girls. Using this method they found evidence of comprehension and production of gender labels in 26-month-olds. Thompson (1975) presented 2- to 3-year-old children with pairs of pictures of male and female adults and children and asked the children to touch the picture indicated by the experimenter. Thompson used many different gender labels and pronouns including man-woman, mommy-daddy, boy-girl, and he-she. He found that using this method, which did not require verbal answers, children as young as 24 months were able to distinguish between the sexes. Thompson did not, however, provide separate analyses for the different word pairs used, and neither the Weinraub et al. nor the Thompson study analyzed results separately for the adult and child pairs.

Leinbach and Fagot (1986) studied a group of 16- to 42-month-old children. They used a method similar to that of Thompson (1975), but they looked separately at toddlers' ability to label adults and children and chose only one pair of gender labels for each; mommy-daddy and boy-girl. They used books of 12 pairs of pictures taken from mail order catalogues for each test. The tests were conducted on a pass/fail basis. A child was required to be correct on at least ten of the twelve pairs in order for performance to be above chance level. Forty-five percent of the children below

the median age (25.7 mos.) passed the adult test, as did 97% of those above the median age. In fact, over 30% of those under 23 months of age passed the adult test. The child test was mastered significantly later than the adult test, but even this test was passed by some as young as 24 months of age. These findings indicate that some infants under two years of age are aware of at least one pair of adult gender labels, and that they are capable of matching those labels with appropriate pictures.

Etaugh, Grinnell and Etaugh (1989) examined the ability of 20-to 42-month-olds to discriminate between male and female pictures of preschoolers, young elementary school children, and young adolescents. They used a pointing method similar to that of Leinbach and Fagot (1986), but used official school photographs which allowed them to control for the age of the children in the pictures and the pose. The age of the children in the pictures did not affect the toddlers' performances. There was a sharp transition from chance to accurate responding at approximately 29 months of age. Regardless of whether they were looking at pictures of preschoolers, school-age children or young adolescents, toddlers under 29 months of age performed at chance levels, whereas those 29 months of age and above performed quite accurately. These findings indicate that accurate labeling of the gender of children is commonly established by two and a half years of age.

To summarize, by using the method of having toddlers point at pictures in response to labels, discrimination of male and female adults using labels has been demonstrated in some toddlers before 23 months of age. Some toddlers have also been shown to be able to discriminate the gender of children by 24 months of age, and most toddlers are proficient at this by 2.5 years of age. All of these findings are well below the three years of age at which traditional gender identity tests are usually passed.

While asking children to point at pictures appears to be an improvement over requiring them to express their knowledge verbally, there is some indication that it is still quite difficult for some toddler-age children. Leinbach and Fagot (1986) reported that 32% of the boys, and 17% of the girls in their study were unwilling or unable to perform their task. All of the children who were unable to participate were below the median age, and therefore ranged between 16 and 26 months of age. Perhaps the requirement of pointing, independent of their ability to discriminate between the sexes using labels, is posing difficulty to many of the subjects under two years of age. Murphy (1978), studied the pointing behaviour of 9-, 14-, 20-, and 24-month-old infants while they were looking at story books with their mothers. While pointing was evident in some 9-month-olds and most 14-month-olds, Murphy observed that these infants appeared to be acting upon the books rather than looking at them. The 20- and 24-month-olds pointed much more

frequently than the 9- and 14-month-olds. Beginning at 20 months the infants' produced verbal labels for some of the objects they pointed at. This study showed that, over time, infants' use of gestures, such as pointing, gradually becomes more sophisticated. Murphy also indicated that pointing ability varies greatly according to the situation. It is likely that pointing in response to a given command, as is required by the gender labeling task, is somewhat more difficult than pointing on one's own initiative. Since pointing behaviour seems to become well established at around 2 years of age, it is not surprising that some subjects under 2 years of age have difficulty with tasks that require pointing.

Although much is now known about the abilities of children over 2 years of age, one of the major difficulties in investigating the early development of knowledge about gender has been finding a suitable method for testing the knowledge of preverbal infants and young toddlers. It is often very difficult for infants and young toddlers to demonstrate their knowledge clearly because of their difficulty with expressing themselves verbally, with responding consistently and with understanding and following directions. Because of these methodological difficulties, little is currently known about the origins of gender related knowledge before two years of age. There is, however, some evidence, from research areas other than those directly examining the ability to identify

and label the sexes, that indicates that infants may be able to distinguish males from females long before they have the linguistic ability to apply gender labels to themselves and others.

Indirect evidence of the ability to distinguish men from women can be inferred from stranger approach studies. In these studies it has been found that young infants respond more negatively to men than to women (Greenberg, Hillman & Grice, 1973; Lewis & Brooks, 1974; Morgan & Ricciuti, 1969). It is likely that these young infants are reacting to gender attributes such as body size, voice pitch, and hair length. Although it is likely that these reactions are based mostly on perceptual cues, these findings indicate that these young infants are capable of differentiating between characteristics of men and women. Studies of early lexical development also provide indirect evidence that children under 2 years of age can distinguish males from females. In these studies it has been found that when overextending social labels such as "daddy", infants as young as 15 months rarely apply such labels to someone of the inappropriate gender (Brooks-Gunn & Lewis, 1979; Thomson & Chapman, 1977).

Recently, studies have been conducted that attempt to directly investigate the acquisition of gender knowledge in preverbal infants. These studies have used habituation paradigms and familiarization-novelty procedures. These paradigms rely on the fact that, from around 2 months of age,

infants have a preference for novel stimuli (Fagan, Fantz, & Miranda, 1971). For example, 7-month-old infants have been shown to habituate to a series of faces of one gender, to generalize that habituation to a novel face of the same gender but to dishabituate when presented with a face of the other gender (Cohen & Strauss, 1979). Since infants prefer to look at novel stimuli, their dishabituation to a face of the other gender indicates that they recognize that it is different in some way from the series of pictures of the first gender. Using familiarization-novelty procedures investigators have shown that infants can distinguish male from female faces (Cornell, 1974; Fagan, 1976; Fagan & Singer, 1979) and male from female voices (Miller, 1983; Miller, Younger & Morse, 1982) by 7 months of age. A study by Leinbach (1990) however, demonstrated that performance on these tasks seems to be based on fairly superficial cues such as hair length and clothing cues. For example, Leinbach found that if infants habituated to a series of pictures of men, and were then presented with a picture of a woman, they would dishabituate as expected. However, if the woman in the picture had her hair shortened to the same length as the men's hair a significantly smaller proportion of the infants dishabituated.

Another difficulty with the familiarization-novelty procedure is that it is possible for infants participating in these studies to form ad hoc categories based entirely on the familiarization stimuli (Reznick, 1989). Using these

procedures, it is not possible to determine whether infants are responding based on pre-existing category knowledge, or whether they are forming new ad hoc categories during the procedure. Data from these studies, therefore, indicate that young infants are capable of distinguishing between male and female voices and faces based on perceptual features but do not provide conclusive evidence that they do so naturally at that age.

The preference looking paradigm has helped to clarify this issue. The preference looking paradigm has been an extremely useful research tool in investigating infant perception and cognition (Spelke, 1985). In this paradigm, infants are typically presented with two visual displays side by side and one auditory sound track presented from the mid point between the visual displays. The sound track corresponds with only one of the visual displays. Infants typically prefer to look at the visual display that corresponds with the sound track. If the infants are capable of determining which visual display corresponds to the sound track, they will, therefore, spend more time looking at that display than at the display that does not correspond to the sound track. In this paradigm it is not possible for infants to form ad hoc categories based on the stimuli presented. The infants are not reinforced for looking at the appropriate screen and there is no opportunity for them to learn which display corresponds to which sound track during the procedure.

If they show a preference for the screen that matches the sound track, then that preference must be based on their pre-existing knowledge of the relationship between the auditory and visual stimuli that were presented.

This paradigm has been used to examine infants' detection of invariant temporal structure and common spatial location (Bahrick, Walker & Neisser, 1981; Spelke, 1976, 1979). For example, Spelke (1976) presented infants with two visual displays, one of a woman playing peek-a-boo and one of a hand repeatedly striking a drum and tambourine. Each infant saw the films twice, once with each sound track. The infants looked primarily at the film specified by the sound track. From this study, Spelke (1976) was able to conclude that infants are able to perceive relations between sights and sounds in the absence of spatial cues.

Spelke and Owsley (1979) used a similar procedure to demonstrate that infants tend to look more at their mother when they hear her voice and more at their father when they hear his voice. Evidence of this ability was found as early as 3.5 months of age, and this effect was quite strong by 7.5 months of age. This study demonstrates intermodal knowledge about individual familiar males and females.

Some understanding of gender as a category system would be demonstrated if it could be shown that infants match the auditory and visual cues associated with general categories of male and female voices and faces. In order to match the

auditory with the visual cues it would be necessary for infants to collect these features in gender differentiated categories. Although largely at the perceptual level, this ability to match the voices and faces would, therefore, indicate that infants have developed rudimentary gender categories.

Walker-Andrews, Bahrick, Raglioni and Diaz (1991) used the preference looking paradigm to investigate this ability. In one study they presented videotapes on which both a man and woman were displayed, both continuously speaking. The sound of only one of the voices was presented at a given time. Temporal synchrony between both of the faces and the voice was controlled. The infants were presented with two 2-minute trials. The authors found that overall 6-month-olds significantly increased their looking to a face when the gender-appropriate voice was played. When the trials were examined separately, however, this effect was only significant for the second trial. When subjects were examined individually, 12 of the 16 6-month-olds spent more time looking at the speaker whose voice was being played than at the other speaker. Results for a group of 4-month-olds were less clear, although it did appear that they could perform the task on at least one trial. In a second study, pairs of pictures of men and women speaking were presented on two colour monitors. The infants were presented with 16 20 second trials, broken into two blocks of 8 trials. As a group 6.5-

month-olds were able to match the auditory and visual cues during the first block of trials but not the second. Neither of these studies found any significant differences between the performances of girls and boys. These two studies seem to indicate that infants as young as 6 months of age may be capable of matching auditory and visual gender cues.

Another recent study (Poulin-Dubois, Serbin, Kenyon & Derbyshire, under review) investigated the ability of 9- and 12-month-olds to match auditory and visual gender cues using the preference looking paradigm. The infants were presented with computer digitized pictures of men's and women's faces on computer monitors while they were seated facing the monitors. The pictures were still pictures and therefore, the mouths were not moving. Male and female voices called out to the infants from a speaker located at the midpoint between the monitors. If the infant spent more time looking at the male pictures when the pictures were presented with a male voice and more time looking at the female pictures when the pictures were presented with a female voice then it was inferred that he or she was capable of matching the auditory with the visual cues. It was found that almost half (45-50%) of the 9-month-olds were able to match the auditory and visual cues and that 62% of the 12-month-olds were able to do so. As in the Walker-Andrews et al. (1991) study, no differences were found between the performances of the boys and girls.

It is likely that the use of black and white, still photographs made this task somewhat more difficult for the infants in the Poulin-Dubois et al. (under review) study. Many of the 9-month-olds in that study were unable to match the voices and faces when, with the use of more colourful and dynamic stimuli, Walker-Andrews et al. (1991) found evidence of the ability in 6-month-olds. Nevertheless, even with this seemingly more difficult task, evidence of rudimentary gender categories was found at a significantly younger age than current theories would predict.

The use of the preference looking paradigm in investigating the development of gender concepts has helped to reduce extraneous task demands. Through the use of this paradigm evidence of preliminary gender categories has been demonstrated at ages where it is unlikely that infants are able to understand gender labels. In 1985, Beverly Fagot argued that sophisticated knowledge about gender might not be necessary for sex-typed behaviour and attitudes to develop, and that perhaps merely an ability to label the sexes is sufficient for children to start forming rules about gender. These recent studies provide a hint that infants may start organizing knowledge about gender even before they can label the sexes. The relationship between the ability to match male and female voices and faces and the ability to understand gender labels has not, however, been examined directly. Such a study would allow for an examination of patterns of

acquisition of these two abilities, and would hopefully provide direct evidence that the development of rudimentary gender categories precedes the ability to understand gender labels.

These two abilities could both be examined using the preference looking paradigm. The preference looking paradigm has been used to investigate infant word comprehension (Golinkoff, Hirsh-Pasek, Cauley & Gordon, 1987; Reznick, 1990). In these studies infants are presented with pictures of two objects and asked to look at one of them. A similar procedure could be used to investigate comprehension of gender labels.

In the present study the preference looking paradigm was used to examine the ability of one group of 18-month-old infants to match male and female voices and faces, and their ability to match the labels 'lady' and 'man' with pictures of men's and women's faces. In choosing the age of the subjects for this study available literature on the comprehension and production of gender labels was considered. Previous studies cited above have not found consistent evidence of comprehension of gender labels much before 2 years of age. Weinraub et al. (1984) found evidence of both comprehension and production of gender labels in 26-month-olds. Data from a longitudinal study of early lexical development (Poulin-Dubois, Personal Communication, May, 1991) has shown that a group of 9 subjects first produced the word man at a mean age

of 21 months, with a standard deviation of 3 months. These subjects first produced the word lady at 20 months of age with a standard deviation of almost 4 months. Since comprehension of words typically precedes production by up to about 5 months, it seems likely that many infants first understand these words in the second half of their second year. As comprehension of gender labels is a linguistic skill, and there is evidence that girls develop language more quickly than boys before 2 years of age (Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991), it is also likely that more girls than boys understand these words at 18 months of age.

Hypotheses

Hypothesis 1: As a group, 18-month-old infants will show a clear ability to match male and female voices with male and female faces.

> Prediction 1: When examined individually almost all subjects will show the ability to match male and female voices and faces.

Hypothesis 2: Evidence that the ability to understand gender labels is emerging will be found.

Prediction 1: It is unlikely that as a group 18-month-olds will show an ability to match the gender labels 'lady' and 'man' with male and female faces.

Prediction 2: When examined individually some of these infants will be able to match the gender labels with male and female faces.

Prediction 3: It may be found that more girls than boys show this ability at 18 months of age.

Hypothesis 3: The ability to match male and female voices and faces precedes the ability to understand gender labels.

Prediction 1: Many subjects will show evidence of being able to match the voices and faces but not be able to understand the gender labels.

Prediction 2: Either none or very few of the subjects will show an ability to understand the labels but not an ability to match the voices and faces.

METHOD

Subjects

Subjects were 17 male and 17 female 18-month-old infants. The mean age was 18.18 months (Range=17.5 mos.-18.75 mos.). An additional six subjects participated in the study but were eliminated due to inattention (see criteria for subject elimination below). Subjects were recruited from birth lists provided by the Conseil de la Santé et des Services Sociaux de Montréal Métropolitain and from birth announcements in the Montreal Gazette. All subjects were from predominantly English speaking families.

Stimuli

Four pairs of black and white photographs of men's and women's faces were chosen from among those used in the Poulin-Dubois et al. (under review) study described above. All of the photographs were originally obtained from a modelling agency. These photographs had all been rated as highly masculine or feminine by a group of undergraduate students for the Poulin-Dubois et al. (under review) study. The photographs were computer digitized and were presented on two Apple II GS colour computer monitors. Although the faces varied in size, facial expression, hair colour and intensity of smile, the male and female photograph of each pair were

matched so that these features would be similar within each pair. The digitized faces were all approximately 6 inches by 8 inches in size.

Computer digitized voice recordings were also presented. In one condition (matching), subjects were required to match male and female voices with male and female pictures. The infant heard a male or female voice saying, "Hi baby, look at me. Here I am. Look at me". The voices used were speaking "motherese" to help keep the infants interested. These voices were chosen from among a series of voice samples that had been obtained from undergraduate volunteers and used in the Poulin-Dubois et al. (under review) study. At that time, the voices chosen for this study were all rated by undergraduate students as highly recognizable as male or female.

In another condition (labeling), subjects were required to match the gender labels 'lady' and 'man' with the male and female pictures. Subjects heard a gender neutral voice saying, "Where's the lady/man? Look at the lady/man". The neutral voice was obtained by recording the voice of an adult woman (approximately 30 years of age) directly onto a Macintosh IISi computer, using a built-in digitizer, and then lowering the pitch of the voice. The voice pitch was lowered to .9439 times its original pitch using a sound editing program (Sound Edit). The voice was played back at the lowered pitch and 12 adults, who did not know the true gender

of the speaker, were asked to decide whether the voice was male, female or whether it could be either. Half of the raters said that the voice could be either male or female, and of the remaining raters half said it was male and half said it was female.

Design

On each trial the face of the gender that matched with the verbal instructions was considered to be the 'target' picture. Thus, for example, in the matching condition, the target picture was the female face when the female voice said "Hi baby, look at me" and in the labeling condition, the target was, the female face when the neutral voice said "Look at the lady". All of the voices were presented at the same, pre-set, computer controlled volume.

No-voice trials, consisting of each pair of pictures presented four times on alternating sides, were presented to provide baseline looking times. The no-voice trials were also used to determine if there were side preferences or sex of picture preferences. These trials were alternated with the matching and labeling test trials.

Randomization Procedure: The picture pairs were randomly assigned to a presentation order. For each trial the side on which the male and female of each pair appeared, and which of the two pictures was targeted, was counterbalanced to meet the

following criteria: 1) In both the matching and the labeling conditions, male and female pictures appeared equally often on the left and on the right 2) In each condition, male and female pictures were targeted an equal number of times on the right and left sides and 3) Each pair of faces appeared twice in each condition, once with the male on the right and female on the left, and once with the male on the left and female on the right. The randomization procedure, therefore, ensured that all four picture pairs appeared equally in all possible positions (left/right, target/non-target) overall and within each condition.

Presentation Sequence: There were two presentation sequences. Half of the infants saw the trials in the original order and the other half saw them in reverse order. The test trials (matching and labeling) were grouped into four blocks. Two of the blocks consisted of matching trials alternating with no-voice trials (blocks M1 & M2). The other two blocks consisted of labeling trials alternating with no-voice trials (blocks L1 & L2). Half of the infants saw the blocks in the order M1, L1, M2, L2. The other half saw the blocks in the order L2, M2, L1, M1 with the trials in the reverse order. Half of the subjects, therefore, saw matching trials alternating with no-voice trials first and half saw labeling trials alternating with no-voice trials first. Appendix A outlines the presentation sequences.

Apparatus

Subjects were seated in a portable infant seat that was securely attached to a table. One of the infant's caretakers sat in a chair that was situated directly behind the infant seat. The table at which the infant was seated was centered 3.5 feet from the back wall of a three sided black cubicle (see Figure 1 for a diagram of the apparatus). The back wall of the cubicle was 6 feet tall and 6 feet, 4 inches wide. The walls were also 6 feet tall and were 6.5 feet long. The walls were angled away from the back wall, so that at the position where the infant was seated they were 9 feet, 10 inches apart. On the back panel the infants could see two computer screens placed two feet apart and a blue light bulb used to draw attention to the center of the display. Also located on the back panel was an audio speaker and the lens of a video camera which was used to record the infant's eye movements. Both were centered between the monitors. The speaker was concealed by a screen. The speaker was located 3.5 feet above the floor. The camera lens was 6 inches above the speaker and the blue light bulb was 6 inches above the camera lens. The equipment used to run the experiment was located behind the back panel of the cubicle. This equipment included the two Apple II GS colour monitors, two Apple II GS computers, the speaker, and a Sony 8mm video camera. The entire presentation was controlled by a custom designed computer program. This program permitted the presentation to be paused if an infant

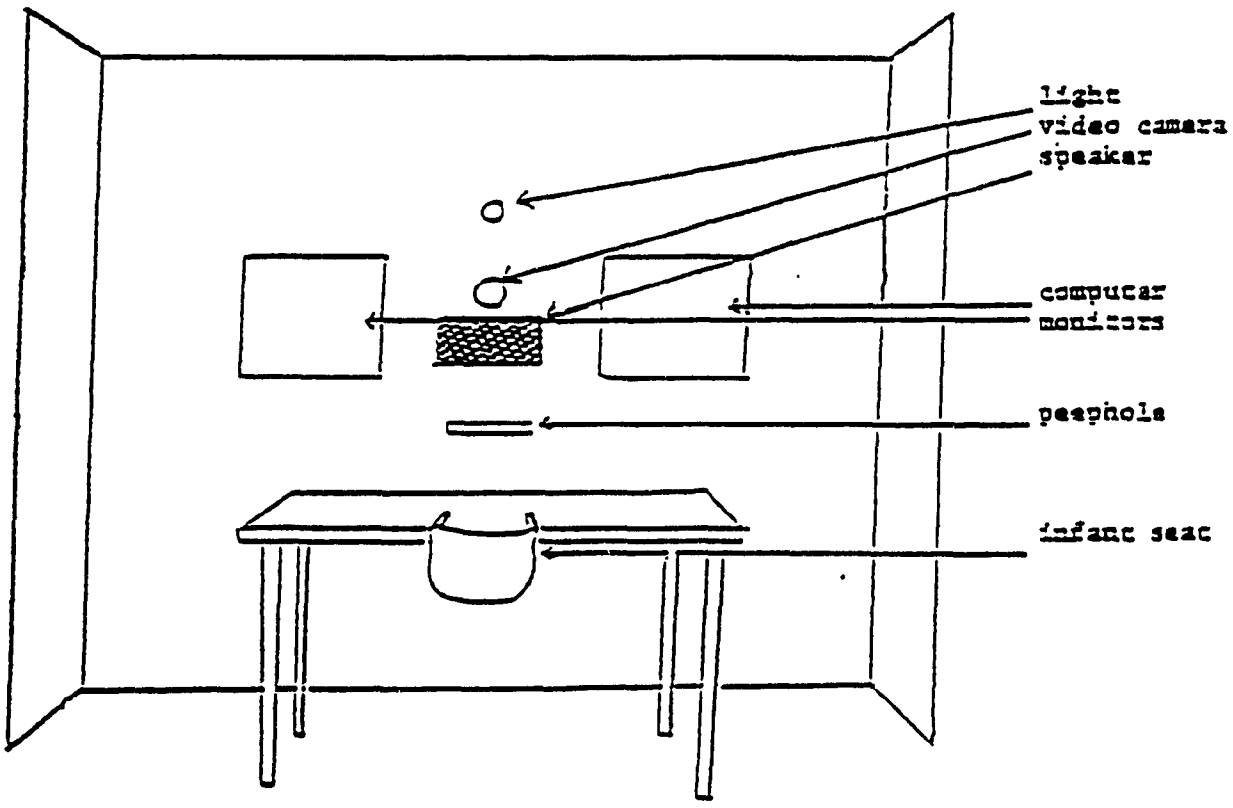

Figure 1 Diagram of the apparatus used for testing subjects

began to fuss or became upset. The videotapes were coded on a MacIntosh computer using a custom designed coding program (Events, Ground Zero Software).

Procedure

The infant and caretaker were greeted and the procedure was explained to the caretaker. The infant and caretaker were then seated in their respective seats. The caretaker was instructed not to touch or talk to their infant while the stimuli were being presented to avoid cueing the infant. He or she was, however, permitted to reorient and reassure the infant during the intertrial intervals.

Each subject saw 34 5-second trials. These consisted of 2 warm-up, practice trials, 8 matching trials, 8 labeling trials, and 16 no-voice trials. Each trial was followed by a 5 second intertrial interval. The beginning of each trial was signalled by the blue light flashing on. The first trial of each half was a warm-up practice trial. On the warm-up trials the infants saw pictures of a dog and a chair and were asked to look at dog. On the voiced trials, the subjects heard the first part of the phrase ("Hi baby, look at me." or "Where's the man/lady?"). The pictures were then presented on the screens and then the infant heard the second part of the appropriate phrase ("Here I am, look at me." or "Look at the lady/man."). On the no-voice trials the procedure was identical except that no voices were presented. If an

infant's fussing resulted in him or her not looking at either screen throughout two consecutive trials the presentation was paused. Presentation resumed when the infant was attentive. The pause command was necessary for approximately one-third of the infants. It was rarely used more than once for any one infant, and the pauses generally lasted from 1 to 5 minutes. The presentation of the trials in each half of the study was continuous unless the pause command was used. All infants were given a short break half way through the testing session. On average the entire test session took approximately ten minutes.

Measures

The amount of time a subject's gaze was directed towards each screen on each trial was coded from the videotapes. The coder could only see a videotape of the infant's eye movements. He or she did not know which side was the "target" side on any given trial. The dependent variable was the average proportion of time the subjects' gaze was directed towards the pictures of each sex when they were targeted compared with the average proportion of time their gaze was directed towards pictures of the same sex on the control trials.

Inter-Observer agreement

In previous similar studies using this coding program, reliability between observers has consistently been found to be around .90. The primary investigator, who was experienced in using the coding program, coded all of the videotapes for the current study. A research assistant then randomly chose 15% of the subjects and coded all of their data to check for reliability. An Intraclass Correlation Coefficient (Ebel, Haertzko, & Carpenter, 1976) was computed between the coders' ratings of amount of time the infants gazed towards each screen on each trial. The Intraclass correlation between coders' ratings across all trials together was .89, $F(319, 320) = 17.12$, $p<.05$.

Results

Subject and Trial Elimination

Trial Elimination Criteria: If an infant did not gaze towards both screens during a trial, and/or if his or her total time gazing at both screens for a trial was less than 1.25 seconds, then that trial was eliminated. If an infant did not gaze towards both screens, then it is not possible to say that he or she preferred the stimulus presented on the one screen that he or she gazed at over the stimulus on the other screen. The less time an infant spent gazing at the screens the less reliable the sampling of the infant's preference was. If the infant's total time gazing at the screens was less than one-quarter of the trial time (< 1.25 seconds) then it is highly likely that the preference sample was unreliable.

Subject Elimination: Subjects were eliminated if more than half of their trials in any condition had to be eliminated due to inattentiveness (as defined above). Subjects were, therefore, eliminated if they did not retain at least half of the target trials for matching, at least half of the target trials for labeling, and at least half of the no-voice control trials. Six subjects were eliminated from the final analyses for this reason.

After trial elimination was completed, and subjects who did not retain enough trials in each condition were eliminated, the data from the remaining subjects was analyzed to check for side bias. Subjects were said to have a side bias if they spent more than 65% of their total looking time across all trials gazing towards the same screen. The value of 65% was arbitrary but it has been used in previous research (e.g. Golinkoff et al., 1987). No subjects were eliminated due to side bias.

Trial Elimination Data for the Final Sample: The mean percentage of trials that were lost for each condition was calculated for the final sample of 34 subjects. On the Matching task subjects lost an average of 13.6% of their target trials and 19.5% of their no-voice trials. On the labeling task subjects lost an average of 12.9% of their target trials and 23.9% of their no-voice trials.

Overview of Analyses

Once subject and trial elimination was completed the data from the final sample were screened to ensure that they did not violate any of the assumptions of Analysis of Variance. Two separate ANOVAs were then conducted, one to examine the effects for the Matching task and one to examine the effects for the Labeling task. Post-hoc simple effects analyses were conducted to examine any interactions that were found. An

alpha level of .05 was chosen as the criterion for statistical significance. Individual differences data were also examined. A pass/fail criterion was established, and the infants' patterns of performance on the two tasks were examined.

Matching Task

The prediction that subjects would be able to match male and female voices with male and female faces (i.e. that they would spend more time gazing towards the pictures of each sex when they were presented with a corresponding voice than when they were presented without a voice) was examined using a 2 (Condition: Target vs. Control trials) x 2 (Sex of Picture) x 2 (Sex of Subject) x 2 (Order of Presentation) mixed-model ANOVA. Condition and Sex of Picture were within subject factors, with Subject Sex and Order as the between group factors (See Table 1 for the means and standard deviations and Appendix B for the complete ANOVA source table). The ANOVA was conducted using data from the 8 target trials for matching and the 8 non-voiced control trials that were alternated with them. Subjects who had not previously been eliminated were excluded from this analysis if they did not have at least half of these 8 non-voiced trials. These subjects were retained during subject elimination because overall, across conditions, they retained at least half of the 16 no-voice trials. One male subject in order 2 was excluded from this analysis for that reason. The results from this ANOVA revealed that, as

TABLE 1

Means and standard deviations for proportions of looking time on the matching trials (A= Male Faces, B= Female Faces)

A	**N**	**Male Targets**		**Male No Voice**	
Group		Mean	SD	Mean	SD
Male Subjects					
order 1	8	.603	.109	.476	.070
order 2	8	.470	.097	.496	.080
Female Subjects					
order 1	9	.535	.120	.443	.047
order 2	8	.536	.105	.455	.039
Entire sample	33	.536	.114	.467	.062

B	**N**	**Female Targets**		**Female No Voice**	
Group		Mean	SD	Mean	SD
Male Subjects					
order 1	8	.565	.044	.524	.070
order 2	8	.548	.125	.504	.080
Female Subjects					
order 1	9	.566	.090	.557	.047
order 2	8	.541	.081	.545	.039
Entire sample	33	.556	.086	.533	.062

N.B. One male subject in order 2 was eliminated from the group analyses for matching as he was inattentive on more than half of the no-voice trials during the matching blocks.

expected, there was a main effect for Condition, $F(1, 29) = 14.23$, $p<.001$. The infants gazed towards the pictures more when they were presented with a voice of the same gender than when they were presented without a voice. The average proportion of time spent gazing towards the pictures when they were targeted was .55, with a standard deviation of .071 which was significantly different from the mean of .50 for the no-voice control trials.

There was also a main effect for Sex of Picture, $F(1, 29) = 7.27$, $p<.05$. Overall subjects spent 52% of their time gazing towards the female pictures ($M = .52$, $SD = .05$) and 48% of their time gazing towards the male pictures ($M = .48$, $SD = .05$). This result indicates that subjects spent more time than would be expected by chance gazing towards the female pictures, and therefore, less time than would be expected by chance gazing towards the male pictures. This preference for the female pictures did not interact with the Condition effect.

There were no significant main effects or interactions with Subject Sex or Order of presentation.

Gender Labeling Task

To examine the prediction that subjects would show evidence of understanding gender labels a separate ANOVA with identical factors to those examined in the ANOVA for matching was used. That is, a 2 (Condition: Target vs. Control trials)

x 2 (Sex of Picture) x 2 (Sex of Subject) x 2 (Order of Presentatic.. xed-model ANOVA with Condition and Sex of Picture as within subject factors, and Subject Sex and Order as the between group factors was conducted. Data from the 8 target trials testing for comprehension of the labels and the 8 non-voiced control trials that were alternated with them was used (See Table 2 for the means and standard deviations and Appendix C for the complete Anova source table). Subjects who had not previously been eliminated were excluded from this analysis if they did not have at least half of these 8 non-voiced trials. One female subject in order 1 was excluded from this analysis for that reason.

The results from this ANOVA revealed that there was an interaction between Condition and Subject Sex, $\underline{F}(1, 29) = 7.18$, $\underline{p}<.05$. A simple effects analysis was conducted to examine this interaction. The girls spent significantly more time gazing towards the faces when they were targeted ($\underline{M} = .54$, $\underline{SD} = .062$) than when they were presented without a voice ($\underline{M} = .50$, $\underline{SD} = .00$), $\underline{F}(1, 15) = 6.84$, $\underline{p}<.05$. In contrast, the amount of time the boys spent gazing towards the faces when they were targeted ($\underline{M} = .47$, $\underline{SD} = .081$) and the amount of time they spent gazing towards the faces on the control trials ($\underline{M} = .50$, $\underline{SD} = .00$) did not differ significantly, $\underline{F}(1, 16) = 1.74$. As a group, therefore, the girls spent more time gazing towards the female faces when asked to "Look at the Lady" and more time gazing towards the male faces when asked to "Look at

TABLE 2

Means and standard deviations for proportions of looking time on the labeling trials (A= Male Faces, B= Female Faces)

A	**N**	**Male Targets**		**Male No Voice**	
Group		Mean	SD	Mean	SD
Male Subjects					
order 1	8	.486	.087	.458	.102
order 2	9	.496	.144	.519	.070
Female Subjects					
order 1	8	.493	.082	.443	.099
order 2	8	.504	.098	.478	.065
Entire sample	33	.495	.102	.476	.086

B	**N**	**Female Targets**		**Female No Voice**	
Group		Mean	SD	Mean	SD
Male Subjects					
order 1	8	.488	.120	.542	.102
order 2	9	.429	.093	.481	.070
Female Subjects					
order 1	8	.651	.089	.557	.099
order 2	8	.515	.061	.522	.065
Entire sample	33	.518	.121	.524	.086

N.B. One female subject in order 1 was eliminated from the group analyses for labeling as she was inattentive on more than half of the no-voice trials during the Labeling blocks.

the Man" indicating that they understood the labels. The boys did not show evidence of this ability.

An interaction between Sex of Subject and Sex of Picture was also found, $F(1, 29) = 4.23$, $p<.05$. A simple effects analysis was conducted to examine this interaction. The girls spent significantly more time gazing towards the female pictures ($M = .54$, $SD = .056$), and significantly less time gazing towards the male pictures ($M = .46$, $SD = .056$) than would be expected by chance, $F(1, 15) = 8.69$, $p<.01$. The times the boys spent gazing towards the pictures of each sex did not differ significantly ($M=.51$, $SD=.07$ and $M=.49$, $SD=.07$ for the male and female pictures respectively), $F(1, 16) = .23$, n.s.. The girls seem to have had a preference for the female pictures whereas the boys did not.

An interaction was also found between Order of Presentation and Sex of Picture, $F(1, 29) = 5.87$, $p<.05$. The simple effects analysis showed that subjects in Order 1 spent more time gazing towards the female pictures ($M = .54$, $SD = .056$), and less time gazing towards the male pictures ($M = .46$, $SD = .056$) than would be expected by chance, $F(1, 15) = 9.53$, $p<.01$. Subjects in Order 2 gazed equally towards the pictures of each sex ($M = .51$, $SD = .069$ and $M = .49$, $SD = .069$ for the male and female pictures respectively), $F(1, 16) = .34$, n.s..

Summary of Group Findings

To summarize the results from the group analyses, it was found that as a group this 18-month-old sample was able to match male and female voices with male and female faces. During the Matching trials the infants also showed a preference for the female faces. This preference did not interact with the ability to match the voices and faces.

As a group the girls were able to match the gender labels 'lady' and 'man' with male and female faces. As a group the boys were not able to do this. During the Labeling trials the girls spent more time gazing towards the female pictures than would be expected by chance, as did subjects in order 1. These preferences for the female pictures did not interact with ability to match the gender labels with the faces.

Individual Differences

In order to examine the hypothesis that infants would show evidence of being able to match the male and female voices and faces before being able to understand the gender labels, pass/fail scores were assigned to each infant for each task. An infant was said to have passed a task if he or she spent more than 50% of his or her looking time looking at the target screen on at least 60% of his or her trials. We expected that the infants' performances would fall into one of three patterns. They would not show an ability to do either task, Pattern 1, they would show evidence of being able to

match the male and female voices with the male and female faces but not be able to understand the gender labels, Pattern 2, or they would be able to do both tasks, Pattern 3. We did not expect the infants to show evidence of understanding the gender labels before being able to match the male and female voices and faces, Pattern 4 (See Table 3 for a breakdown of how subjects fell into these patterns).

Overall, 31 of the 34 subjects (92.8%) fell into one of the 3 expected patterns. Approximately one-quarter of the infants (23.6% of the girls and 29.5% of the boys) did not show an ability to do either task (Pattern 1). Almost two-thirds (64.7%) fell into either Pattern 2 or Pattern 3 and, therefore, showed evidence of being able to match male and female voices and faces. Half of the girls who were able to match the voices and faces were also able to understand the gender labels, whereas, only one-quarter of the boys who could match the voices and faces could understand the labels. Overall, 23.5% of the subjects fell into Pattern 3, showing evidence of being able to do both tasks. Only 3 of the 34 infants (8.8%) showed evidence of being able to understand the gender labels while not being able to match the male and female voices and faces (Pattern 4). All 3 of these infants were girls.

A McNemar test for the significance of changes was calculated for the individual data of the total sample. It was found that it was more likely for infants to be able to

TABLE 3

Subjects' Abilities to Match Voices and Faces and to Understand Gender Labels

	Match	Label	Girls N=17	Boys N=17	Total Sample
Pattern 1	NO	NO	23.6%	29.5%	26.5%
Pattern 2	YES	NO	29.4%	52.9%	41.2%
Pattern 3	YES	YES	29.4%	17.6%	23.5%
Pattern 4	NO	YES	17.6%	0%	8.8%

match and not label than it was for them to be able to label and not match (X^2 (1, N = 34) = 5.88, p<.05).

Discussion

The results of this study provide support for all of the initial hypotheses. The first hypothesis was that almost all of the 18-month-olds would be able to match the voices and faces. The results of the ANOVA demonstrated that as a group the infants showed a definite ability to do this task. While the hypothesis that almost all infants would show this ability is not entirely accurate, most of the infants (64.5%) were able to do this task. Many others may also be capable of this distinction but were not able to demonstrate their abilities on the preference looking task. Poulin-Dubois et al. (under review) found that 45-50% of 9-month-olds and 62% of 12-month-olds in their study could perform the matching task. The proportions of infants who demonstrated an ability to do this task at 12 and 18 months of age appear to be similar. However, a less stringent criteria was used in the Poulin-Dubois et al. (under review) study. In that study, infants were said to have passed a task if they looked at the target screen more than 50% of the time on more than 50% of their trials, whereas, in the current study, infants were said to have passed a task if they looked at the target screen more than 50% of the time on at least 60% of their trials. Due to the differences in the criteria used it is difficult to determine whether more 18-month-olds than 12-month-olds were actually able to do this task.

The second hypothesis was that comprehension of gender labels is emerging at 18 months of age. It was predicted that as a group the 18-month-olds probably would not be able to understand the gender labels, but that this ability would be demonstrated by some of the subjects. It was also predicted that more girls than boys would show this ability. All of these predictions were supported. In the Anova the ability to understand the gender labels interacted with the sex of the subjects. As a group the girls were able to do this task, whereas the boys were not. In total, 17.6% of the boys and 47% of the girls were able to understand the labels. The fact that more girls than boys were able to understand the gender labels does not mean that girls have more advanced knowledge about gender than boys. In general, prior to 2 years of age, girls tend to acquire language more quickly than boys (Huttenlocher et al., 1991). Since comprehension of gender labels is a linguistic skill, it is not surprising that, in this 18-month-old sample, girls were better able to perform this skill than boys. If one looks at the data from the matching task there is no evidence that the girls perform better than the boys. In fact, if there is any difference at all, the performance of the boys appears to be somewhat superior.

The final hypothesis was that the ability to match voices and faces precedes the ability to understand gender labels. In general, this also appears to be true. Only 3 of the

subjects showed an ability to understand the labels without also showing an ability to match the voices and faces. While some of the infants were able to do both the matching and the labeling tasks, many of the infants (half of the girls and three-quarters of the boys) who showed an ability to match the voices and faces were not able to understand the labels.

The most important finding of this study seems to be that infants appear to acquire intermodal knowledge about gender before they are able to understand gender labels. This is important because it has significant theoretical implications. The fact that infants appear to begin to organize knowledge by gender before they understand gender labels, implies that they can develop gender categories before they have specific labels to attach to those categories. Rather than being a catalyst that allows children to begin to organize and acquire knowledge about gender, the acquisition of gender labels, may simply allow them to add a verbal label to the knowledge that they have already accumulated in preexisting categories. The ability to understand gender labels does not appear to be a good measure of a child's understanding about gender. It appears that this ability may be more reflective of a child's linguistic capacity than his or her knowledge about gender. This hypothesis is supported by the different patterns of performance of the girls and the boys on the matching and labeling tasks. As was described above, the girls outperformed the boys on the labeling (linguistic) task but not

on the matching task. If it is true that the ability to understand gender labels is not a good indication of the extent of knowledge a child possesses about gender, there is little reason to believe that the acquisition of gender labels is necessary for sex-typed attitudes or behaviours to appear. This is probably why sex-typed behaviour and toy choices (Weinraub et al., 1984; Kuhn et al., 1978) and preference for same-sex playmates (Maccoby, 1988) are evident before the acquisition of gender labels. Although the acquisition of gender labels does not appear to be necessary for these sex-typed preferences to appear, there is some evidence that the ability to label the sexes can have some effect on the strength of these preferences. Fagot, Leinbach and Hagan (1986), for example, found that while all children in their sample showed a preference for same-sex playmates, those who could understand gender labels for children spent more time playing with same-sex peers than those who did not show the ability to understand these labels. In discussing these findings, Leinbach and Fagot (1986) hypothesized that "gender labeling signals the point at which children are aware of the discriminations they may already be making tacitly" (p. 665). If the acquisition of gender labels truly is the point at which infants become consciously aware of the gender categories that they possess, then this awareness may allow the infant to actively seek out more information about gender. An increase in the rate of acquisition of knowledge about

gender may, therefore, be seen around the time when the gender labels are acquired.

To summarize, the results of this study support the hypothesis that rudimentary gender categories develop before gender labels are understood. This helps to explain why sex-typed attitudes and behaviours also develop before comprehension of gender labels. Although it does not appear to be a necessary precursor to the development of these behaviours, gender labeling may be an important marker of the beginning of an infant's conscious awareness of gender categories. In this study comprehension of gender labels was demonstrated at a younger age than in previous studies.

In addition to contributing to our understanding about the development of gender knowledge and sex-typed attitudes and behaviours, the results of this study also make an important contribution to the literature involving methods of testing the knowledge of preverbal infants. The positive results of this study, along with those of Walker-Andrews et al. (1991) and Poulin-Dubois et al. (under review), provide support for the idea that, by reducing the extraneous task demands to a minimum, the preference looking paradigm allows us to explore the emerging knowledge of young infants at ages where they are unable to respond consistently to more demanding tasks. Comprehension of gender labels was found in infants almost half a year younger than the youngest age at which this ability has previously been demonstrated.

The results, however, also demonstrate some of the limitations of this paradigm. It seems that although positive results found using this paradigm are generally accurate, the meaning of negative results is not as clear. For example, it seems unlikely that over one-third of the 18-month-old infants that were tested were truly unable to match male and female voices with male and female faces. This would indicate that a large number of 18-month-olds are not capable of a task that has been found to emerge in some infants as young as 9 months of age. In addition, informal observation of infants this age would lead one to expect that they do have the ability to differentiate men from women.

It seems likely that many of those infants who did not show an ability to do either of the tasks were actually capable of making at least one of the distinctions being tested. Although the preference looking paradigm is simpler than many other tasks, quite a few infants are still not able to perform this task. In terms of the current study, it can be assumed that infants who showed an ability to perform a certain task actually possessed the knowledge being tested, whereas infants who did not show an ability to do a given task may or may not have actually possessed the required knowledge. This difficulty with interpreting negative results found using the preference looking paradigm is shared with other infant visual perception tasks such as the habituation tasks and familiarization-novelty procedures described earlier. This

difficulty was evident in Robert Fantz's (1961) original preference method. Fantz presented infants with two visual stimuli and recorded the amount of time they looked at each to determine whether they had a preference for one of the stimuli over the other. If a preference was demonstrated, then one could infer that an infant was able to discriminate between the stimuli. In discussing this paradigm, Miller (1987) indicated that if an infant did not show a preference on this task it could mean that the infant was not able to discriminate between the stimuli, but it was also possible that the infant could make the distinction but simply had no preference. Flavell (1985) summed up the difficulty in interpreting negative results using these paradigms by stating that "preference logically implies discrimination but discrimination certainly does not imply preference" (p. 170).

It is likely that the three infants who showed an ability to understand the gender labels but did not match the male and female voices and faces were actually capable of both tasks. Since we can be quite certain that these infants were able to understand the gender labels, and since it appears that this ability is generally acquired after the ability to match the voices and faces, it could be that the voice and face matching task was well within the capability of these three infants but for some reason they simply did not show a preference for looking at the faces that matched the voices.

In summary, one cannot infer an inability to make a distinction based on negative results found using this method. It would, however, be very difficult for an infant to pass such a task by chance alone. Positive results, in contrast, are therefore probably quite accurate.

Although the preference looking paradigm does seem to underestimate the abilities of some of the young infants being tested, it has been shown to be more sensitive than other methods that are currently available. The continued use of this paradigm in investigating the knowledge possessed by young infants seems to hold a lot of promise. It is likely that through the continued use of this paradigm it will be discovered that young infants possess more knowledge than we have imagined. The use of this paradigm in investigating the acquisition of knowledge about gender is a relatively new development in the area. Since extensive knowledge about gender and about sex-role stereotypes has been found in very young children, it seems likely that the use of this paradigm will allow researchers to discover more about the roots of gender knowledge in infancy.

The fact that this and other recent studies have found that male and female voices can be used effectively to cue infants to look differentially at visual displays holds particular promise. These voice cues could be used to direct infants' looking at visual stimuli other than male and female faces. For example, a study currently in progress is using

boys' and girls' voices to direct infants' looking at pairs of male and female sex-typed toys. In this way knowledge of the sex-typing of toys is being explored in young infants. The types of knowledge that may increase after the acquisition of gender labels could be explored by combining tasks such as the toy preference study described above with the gender labeling task and looking at patterns of abilities. The introduction of the preference looking paradigm to this field of study has made the exploration of numerous new and exciting research avenues possible.

References

Bahrick, L.E., Walker, A.S. & Neisser, U. (1981). Selective looking by infants. Cognitive Psychology, 13, 377-390.

Bartko, J., & Carpenter, W. (1976). On the methods and theory of reliability. The Journal of Nervous and Mental Disease, 163, 305-317.

Brooks-Gunn, J. & Lewis, M. (1979). "Why mama and papa?" The development of social labels. Child Development, 50, 1203-1206.

Cohen, L.B. & Strauss, M.S. (1979). Concept acquisition in the human infant. Child Development, 50, 419-424.

Cornell, E.H. (1974). Infants' discrimination of photographs of faces following redundant presentations. Journal of Experimental Child Psychology, 18, 98-106.

Etaugh, C., Grinnell, K., & Etaugh, A. (1989). Development of gender labeling: Effect of age of pictured children. Sex Roles, 21 (11/12), 769-773.

Fagan, J.F. (1976). Infants' recognition of invariant features of faces. Child Development, 47, 627-638.

Fagan, J.F., Fantz, R.L., & Miranda, S.B. (1971). Infants' attention to novel stimuli as a function of postnatal and conceptual age. Paper presented at the meeting of the society for Research in Child Development, Minneapolis, April.

Fagan, J.F., & Singer, L.T. (1979). The role of simple feature differences in infants' recognition of faces. Infant Behaviour and Development, 2, 39-45.

Fagot, B.I. (1985). Changes in thinking about early sex-role Development. Developmental Review, 5, 83-98.

Fagot, B.I., Leinbach, M.D., & Hagan, R. (1986). Gender labeling and the adoption of sex-typed behaviours. Developmental Psychology, 22, 440-443.

Fantz, R.L. (1961). The origin of form perception. Scientific American, 204, 66-72.

Flavell, J.H. (1985). Cognitive Development (pp. 167-198). Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Golinkoff, R.M., Hirsh-Pasek, K., Cauley, K.M. & Gordon, L. (1987). The eyes have it: lexical and syntactic comprehension in a new paradigm. Journal of Child Language, 14, 23-45.

Greenberg, D.J., Hillman, D. & Grice, D. (1973). Infant and stranger variables related to stranger anxiety in the first year of life. Developmental Psychology, 9 (2), 207-212.

Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M. & Lyons, T. (1991). Early vocabulary growth: Relation to language input and gender. Developmental Psychology, 27 (2), 236-248.

Kohlberg, L.S. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E.E.

Maccoby (ed.). The Development of Sex Differences (pp.82-173). Stanford, CA.: Stanford University Press.

Kuhn, D., Nash, S.C. & Bruchen, L. (1978). Sex-role concepts of two-and three-year-olds. Child Development, 49, 445-451.

Leinbach, M.D. (April, 1990). Infants' use of hair and clothing cues to discriminate pictures of men and women. Paper presented at the Seventh International Conference on Infant Studies, Montreal, Quebec.

Leinbach, M.D. & Fagot, B.I. (1986). Acquisition of gender labels: A test for toddlers. Sex Roles, 15 (11/12), 655-666.

Lewis, M. & Brooks, J. (1974). Self, other, and fear: Infants' reactions to people. In M. Lewis & L.A. Rosenblum (Eds.), The origins of fear: The origins of behaviour (Vol. 2, pp. 195-227). New York: Wiley.

Maccoby, E.E. (1988). Gender as a social category. Developmental Psychology, 24(6), 755-765.

Maccoby, E.E. (1980). Social Development. New York: Harcourt Brace Jovanovich.

Miller, C.L. (1983). Developmental changes in male/female voice classification by infants. Infant Behaviour and Development, 6, 313-330.

Miller, C.L., Younger, B.A., & Morse, P.A. (1982). The categorization of male and female voices in infancy. Infant Behaviour and Development, 5, 143-159.

Miller, S.A. (1987). Developmental Research Methods (pp. 99-107). Englewood Cliffs, N.J.: Prentice Hall, Inc.

Morgan, G.A. & Ricciuti, H.N. (1969). Infants' responses to strangers in the first year. In B.M. Foss (Ed.), Determinants of infant behaviour (Vol. 4, pp. 253-272). London: Methuen.

Murphy, C.M. (1978). Pointing in the context of a shared activity. Child Development, 49, 371-380.

Poulin-Dubois, D., Serbin, L.A., Kenyon, B.L., & Derbyshire, A.J. (Under Review). Infants' Intermodal Knowledge About Gender. Developmental Psychology.

Reznick, J.S. (1989). Research on infant categorization. Seminars in Perinatology, 13 (6), 458-466.

Reznick, J.S. (1990). Visual preference as a test of infant word comprehension. Applied Psycholinguistics, 11, 145-166.

Slaby, R.G. & Frey, K.S. (1975). Development of gender constancy and selective attention to same-sex models. Child Development, 46, 849-856.

Spelke, E.S. (1976). Infants' intermodal perception of events. Cognitive Psychology, 8, 553-560.

Spelke, E.S. (1979). Perceiving bimodally specified events in infancy. Developmental Psychology, 15, 626-636.

Spelke, E.S. (1985). Preferential-looking methods as tools for the study of cognition in infancy. In G. Gottlieb & N.A. Krasnegor (Eds.), Measurement of audition and vision

in the first year of postnatal life. A methodological overview (pp. 323-363). Norwood, NJ: Ablex.

Spelke, E.S. & Owsley, C.J. (1979). Intermodal exploration and knowledge in infancy. Infant Behaviour and Development, 2, 13-27.

Thompson, S.K. (1975). Gender labels and early sex role development. Child Development, 46, 339-347.

Thomson, J.R. & Chapman, R.S. (1977). Who is 'daddy' revisited: The status of two-year-olds' over-extended words in use and comprehension. Journal of Child Language, 4, 359-375.

Walker-Andrews, A.S., Bahrick, L.E., Raglioni, S.S. & Diaz, I. (1991). Infants' bimodal perception of gender. Ecological Psychology, 3 (2), 55-75.

Weinraub, M. & Brown, L.M. (1983). The development of sex-role stereotypes in children: Crushing realities. In V. Franks & E. Rothblum (Eds.), The stereotyping of women: Its effects on mental health. (pp. 30-58). New York: Springer.

Weinraub, M., Clemens, L.P., Sockloff, A., Ethridge, T., Gracely, E., & Meyers, S.B. (1984). The development of sex role stereotypes in the third year: Relationships to gender labeling, gender identity, sex-typed toy preference, and family characteristics. Child Development, 55, 1493-1503.

Appendix A

Presentation Sequences

Appendix A

	Order 1			Order 2	
	Left	Right		Left	Right
M-1	F2	**M2**	L-2	f2	m2
	f4	m4		M2	**F2**
	F1	M1		m1	f1
	m3	f3		F3	**M3**
	F3	**M3**		m3	f3
	m1	f1		**F1**	M1
	M2	**F2**		f4	m4
	f2	m2		F2	**M2**
L-1	**F4**	M4	M-2	m2	f2
	m4	f4		**M4**	F4
	M3	**F3**		f3	m3
	f1	m1		**M1**	F1
	M1	F1		f1	m1
	f3	m3		M3	**F3**
	M4	F4		m4	f4
	m2	f2		**F4**	M4
			Break		
M-2	**F4**	M4	L-1	m2	f2
	m4	f4		**M4**	F4
	M3	**F3**		f3	m3
	f1	m1		**M1**	F1
	M1	F1		f1	m1
	f3	m3		M3	**F3**
	M4	F4		m4	f4
	m2	f2		**F4**	M4
L2	F2	**M2**	M-1	f2	m2
	f4	m4		M2	**F2**
	F1	M1		m1	f1
	m3	f3		F3	**M3**
	F3	**M3**		m3	f3
	m1	f1		**F1**	M1
	M2	**F2**		f4	m4
	f2	m2		F2	**M2**
	Total 32 trials			Total 32 trials	

Bold faced trials are target trials
Lower case trials are no-voice trials
F2 =Female of picture pair 2

M-1 = Matching block 1
L-2 = Labeling block 2

Appendix B

ANOVA Source Table for Matching Trials

Appendix B

ANOVA Source Table for Matching Trials

Independent variables:

voice/no voice (condition/cnd)	-within Ss
sex of picture (picsex)	-within Ss
sex of subject (sbjsex)	-between Ss
order (ord)	-between Ss

Dependent variable: proportion of total time gazing at the screens

SOURCE	SS	df	MS	F	p
Betw Ss					
sbjsex	.00	1	.00	.01	.939
order	.02	1	.02	3.29	.080
sbjsex by ord	.01	1	.01	1.73	.198
within cell	.14	29	.00		
Within Ss					
picsex	.06	1	.06	7.27	.012*
sbjsex by picsex	.01	1	.01	1.29	.266
ord by picsex	.00	1	.00	.04	.838
sbjsex by ord by picsex	.01	1	.01	1.00	.327
within cell	.23	29	.01		
condition	.07	1	.07	14.23	.001*
sbjsex by cnd	.00	1	.00	.01	.939
ord by cnd	.02	1	.02	3.29	.080
sbjsex by ord by condition	.01	1	.01	1.00	.198
within cell	.23	29	.01		
picsex by cnd	.02	1	.02	1.76	.195
sbjsex by picsex by cnd	.01	1	.01	1.19	.285
ord by picsex by condition	.01	1	.01	1.27	.269
sbjsex by ord by picsex by condition	.01	1	.01	1.34	.257
within cell	.28	29	.01		

Appendix C

ANOVA Source Table for Labeling Trials

Appendix C

ANOVA Source Table for Labeling Trials

Independent variables:

voice/no voice (condition/cnd)	-within Ss
sex of picture (picsex)	-within Ss
sex of subject (sbjsex)	-between Ss
order (ord)	-between Ss

Dependent variable: proportion of total time gazing at the screens

SOURCE	SS	df	MS	F	p
Betw Ss					
subjsex	.04	1	.04	7.18	.012*
order	.02	1	.02	3.20	.084
sbjsex by ord	.00	1	.00	.60	.445
within cell	.14	29	.00		
Within Ss					
picsex	.05	1	.05	3.40	.075
sbjsex by picsex	.06	1	.06	4.23	.049*
ord by picsex	.08	1	.08	5.87	.022*
sbjsex by ord by picsex	.00	1	.00	.02	.878
within cell	.42	29	.01		
condition	.00	1	.00	.41	.529
sbjsex by cnd	.04	1	.04	7.18	.012*
ord by cnd	.02	1	.02	3.20	.084
sbjsex by ord by condition	.00	1	.00	.60	.445
within cell	.14	29	.00		
picsex by cnd	.00	1	.00	.49	.490
sbjsex by picsex by condition	.01	1	.01	.77	.386
order by picsex by condition	.00	1	.00	.03	.860
sbjsex by ord by picsex by cnd	.01	1	.01	.88	.356
within cell	.29	29	.01		