

Acquisition of Gender Labels: A Test for Toddlers¹

Mary Driver Leinbach

University of Oregon

Beverly I. Fagot²

University of Oregon and Oregon Social Learning Center

A test of gender discrimination in response to familiar labels was developed and given to 17- to 42-month-old children. A pretest employing pictures of familiar objects was presented first to ensure that subjects could perform a discrimination task, followed by separate gender tests comprised of photographs of stereotypically masculine and feminine children and adults. There were no sex differences in performance for the gender tests, but among the youngest children, more boys than girls could not be tested. Psychometric aspects of the tests were investigated and found adequate. The tests allow individual children to be classified as to gender-labeling ability and provide a useful tool for investigating gender knowledge.

Gender identity, that is, self-categorization as a member of one sex or the other, has been specified as the point at which the process of sex typing begins: The desire to act in accord with self-perceived identity as a boy or girl is thought to motivate the child's acquisition of sex-typical attitudes and behaviors (Kohlberg, 1966; Maccoby, 1980). Children who can label themselves and others accurately are presumed to have attained gender identity, the first of Kohlberg's three stages in the development of secure and

¹This research was funded in part by BRSB Grant S07RR07080 awarded by the Biomedical Support Program, Division of Research Resources, National Institutes of Health. This article is based on a dissertation submitted by the first author in partial fulfillment of the requirements for the doctoral degree at the University of Oregon. A preliminary report of these findings was presented at the biennial meeting of the Society for Research in Child Development, Detroit, 1983.

²To whom reprint requests should be addressed at Department of Psychology, University of Oregon, Eugene, Oregon 97403.

stable gender understanding. Unfortunately, adopting accurate labeling of self and others as the *starting point* of sex role development has deflected empirical attention away from the process of label acquisition, and inadequate understanding of children's earliest gender knowledge has contributed to confusion over relationships between cognitive accomplishment and sex-typed behavior.

In general, sex-typed behavior and attitudes are found earlier than Kohlberg's (1966) cognitive-developmental theory would predict, and appear to precede accurate labeling of self and others (Constantinople, 1979; Huston, 1983; Lewis & Weinraub, 1979). Although the average age at which children can classify themselves and others well enough to meet cognitive-developmental standards for gender identity is approximately three years (Slaby & Frey, 1975), knowledge of sex role stereotypes has been found by 30 months (Gettys & Cann, 1981; Kuhn, Nash, & Brucken, 1978); 26-month-olds have shown gender labeling, gender identity, and awareness of adult sex role differences (Weinraub, Clemens, Sockloff, Ethridge, Gracely, & Meyers, 1984); and sex-typical toy preferences have been found consistently before 24 months (Blakemore, LaRue, & Olejnik, 1979; Fagot, 1974, 1978; Fein, Johnson, Kosson, Stork, & Wasserman, 1975).

It may be that cognitive-developmental theory, in emphasizing attainment of successively higher levels of reasoning, overestimates the degree of understanding children must possess before knowledge of sex typing can begin to affect behavior. Tests of gender knowledge have been constructed to ensure that achievement of a particular level is not credited unless the child can hold fast to a correct judgment of a target child's sex in the face of a perceptual transformation of the stimulus (Emmerich, Goldman, Kirsh, & Sharabany, 1977) or repeated questioning (Slaby & Frey, 1975). The nature of these tasks and stringent criteria for reaching Kohlberg's levels have tended to mask early knowledge and to leave children under three looking curiously incompetent with regard to gender. Even very young children may possess incipient gender categories and a better command of gender labels than they have been given credit for.

Recent investigations indicate that children's knowledge of categories is greater than their typically poor performance on classification tasks would suggest, and that even young children use category labels to support inductive inferences about category members. In particular, children who are unable to classify objects on the basis of shared properties are often quite adept at inferring properties when presented with category labels (Gelman & Markman, 1983). Gelman, Collman, and Maccoby (1986) have extended this finding to early gender knowledge. They showed that preschool children who fail gender constancy tasks, which require that category membership be deduced from information about properties or attributes, can nevertheless

infer sex-typical properties and attributes when given gender labels. That is, being told that someone is a boy or girl allows the child to demonstrate categorical knowledge of what boys and girls are likely to be or to do. Gelman et al. proposed that even very young children recognize gender-category labels as important pieces of information that permit inferences about enduring properties of the category members.

As Kohlberg (1966) pointed out, children generally learn to say "boy" and "girl" at about two years of age, before they know to whom the labels apply. Even before other children can be categorized accurately, the child may be acquiring information about what boys and girls are supposed to look like or to do. When labels can be used to designate self and others correctly, children should be able to use whatever gender knowledge they have to inform their own behavior and interpret that of others around them. However, ascertaining the onset of accurate labeling calls for a testing procedure that takes the toddler's limited verbal ability into account and is sensitive to the beginning of competence.

Children just learning to talk seem to understand much more than they can say (Nelson, 1973), and must be given a simple task and nonverbal response mode if they are to demonstrate what they know. Awareness of this issue led Thompson (1975) to include a nonverbal gender-discrimination task in his widely cited investigation of gender labeling and sex role development in two- to three-year-olds; children were shown male-female pairs of adults and children, and asked to point to the named member of each pair. However, the use of nine different label pairs—including child terms (boy-girl, brother-sister), adult terms (man-woman, man-lady, mommy-daddy, father-mother), and pronouns (he-she, him-her, his-her)—in the 17 trials given make the results somewhat difficult to interpret. Except for noting that the brother-sister pair elicited the most errors, Thompson did not report whether children responded differently to items featuring children and adults, or whether stimulus pictures differed in discriminability. Nor is it clear that the probability of succeeding by chance on these binomial choices was given sufficient consideration. Although this study contributed substantially to methodology for studying gender knowledge in young children, a more adequate test of the child's knowledge of gender labels would deal with the label pairs separately and show greater concern for chance effects.

The present study was designed to develop an appropriate task for investigating knowledge of gender labels in very young children. Such a test must be nonthreatening and easy enough to reveal what knowledge the child may have, simple to administer, and psychometrically sound. In our view, these goals could be best met by showing the child a series of pictures representing the discrimination involved, and assessing the ability to choose the named

member of each pair more consistently than chance responding would predict. The labels originally chosen were "boy" and "girl" because of their importance in Kohlberg's theoretical account of sex typing and because they are the most relevant to children's ability to label themselves. An additional label set was added when pilot work revealed that adults were identified by some children unable to identify boys and girls.

Obtaining normative data on very young children's grasp of gender labels was the second purpose of this study. We expected that children who did not yet apply the labels correctly would respond to the discrimination tasks randomly, and that random responding would be more common in the youngest children. The possibility of sex differences in responding was investigated but not anticipated. The relationship between scores for the child and adult discrimination tests was examined but no specific predictions as to its nature were made.

METHOD

Subjects

Subjects were 122 children (68 boys and 54 girls) between the ages of 16 and 43 months; 103 were members of experimental play groups conducted for the purpose of observing naturally occurring play and peer interaction, and 19 were enrolled in a university day care center. Approximately 15% were children of university students; the remainder were from middle-class and working-class families in the community.

Data for this study were collected over a period of two years, as children of appropriate age became available. Testing was continued until each gender test had been administered to 40 children of each sex, keeping the ages of boys and girls as comparable as possible across the age range of children tested. Three boys (mean age, 25 months) and three girls (mean age, 23.7 months) who had completed a pretest were not available for further testing; 22 boys and 9 girls between the ages of 16 and 26 months could not be tested because they were unable or unwilling to attempt or to complete the pretest. The final sample consisted of 75 children who were given both gender tests – 5 who received only the adult test, and 5 who received only the child test (85 children in all, $n = 80$ for each test). For both tests, subjects' mean age was 26.6 months ($SD = 5.7$, child test, and 5.9, adult test).

Test Materials

Three separate sets of stimulus materials – a pretest and the two gender discrimination tests – were prepared for this study. Each set consisted of 12

picture pairs, individually mounted on white index cards, 5 × 8 inches, presented one pair at a time.

Pretest. The first task consisted of pictures of familiar objects and animals taken from magazines and children's books. To obtain the necessary set of 12 item pairs, 36 pictures were presented for identification to each of ten 18- to 32-month-old children. Sixteen items that were never missed, and eight missed only once, were retained for the pretest. This task was used *only* as a screening device to determine each child's ability to perform a discrimination task.

Child Gender Test. Pictures for the child gender test were selected from 44 colored photographs of boys and girls taken from magazines and mail-order catalogues, each showing only the head and shoulders of a fully clothed child. The sex of each pictured child was judged by 22 university students, 7 men and 15 women; 24 pictures correctly and unanimously identified as boys or girls were retained and arranged as male-female pairs on facing pages of a looseleaf notebook, matched as nearly as possible for size of face and apparent age of the child. Subjects were asked to identify the pictures in response to the words "boy" and "girl."

Adult Gender Test. Twenty-four pictures of men and women, highly stereotypic in their conformity to current standards of clothing, hairstyle, and makeup, were selected from magazines and mail-order catalogues, and prepared and presented as for the child task. "Mommy" and "Daddy" were used as the labels for adults: if a child did not respond to these labels, others such as "man" and "woman" were substituted.

Procedure

The three tasks were administered on separate occasions within one month unless the child was not available for testing, beginning with the pretest. Sequence of presentation of the two gender discrimination tests was counterbalanced. For all three tasks, subjects received one of four permutations of the 12 item pairs in randomly selected order; the choice and position of the member of the picture pair the child was asked to identify were varied systematically. Thus for the gender tests, each male and each female picture was designated as the target on half of the trials, and the target picture occupied the left and right positions equally often.

Children were tested individually in a small room equipped with a low round table partially bisected by a hanging fabric screen. The child was seated upon the table facing the stimulus materials, which were placed on a low stand. One experimenter, seated on a low chair facing the child but unable to see the pictures, named a member of each picture pair and asked the subject to pat, touch, or point to the corresponding picture. A second ex-

perimeter, seated on the table with the child, controlled the stimulus materials. To avoid cuing the child, the first experimenter was blind to the location of the target picture and order of the pairs, and the second experimenter's face was concealed from the first experimenter by the hanging screen during each question and response. Both experimenters talked freely with the child throughout administration of the tests, except that the second experimenter did not discuss the test pictures or questions. The first experimenter repeated or rephrased questions as necessary to secure the child's attention, but no feedback regarding success or failure was given.

Scoring

The criterion for success on the pretest was the ability to respond to the experimenter's request to indicate one member of each item pair; children who could do so were considered testable regardless of whether or not they indicated the correct picture consistently.

The two gender tests were scored as "pass" or "fail," with at least 10 correct choices required to pass. Because the probability of success on any item pair by guessing alone was .50, only scores of 10 or above for each 12-item test could be considered to differ from chance responding beyond the .05 level of significance.

Data for 25 randomly chosen test administrations were recorded from a one-way observation booth, without sound and without the knowledge of the experimenters conducting the tests. Reliability was indicated by percentage of agreement across trials [agreements/(agreements + disagreements)] was 95%.

Data Analysis

Most of the analyses involved dichotomous data, either as individual items correct or incorrect, or as children categorized as passing or failing the tests. As no assumptions could be made regarding the distributions underlying these dichotomies, nonparametric tests were used to assess differences in proportions of children passing each test (older vs younger, boys vs girls), item-order effects, sequence of acquisition of the gender tests, and item difficulty.

RESULTS

Psychometric Aspects of the Gender Tests

Order Effects. To ensure that the order in which the gender tests were presented had not influenced performance on either test, the children who were given both tests were grouped according to test sequence (child test or adult test first), with z tests used to assess differences in proportions of children in each group who passed each test. The possibility of item order effects was investigated for each test by dividing subjects into four groups on the basis the item permutation received and comparing the proportions of children passing in each group. As no differences due to test sequence or item order were found for either test, the groups were combined for the remaining analyses.

Item Analyses. Internal consistency of each gender test was investigated by computing standardized alpha coefficients; for the child test, $\alpha = .624$, and for the adult test, $.728$. Repeated-measures analyses of variance, with items as repeated measures and Cochran's Q as the test statistic, revealed that the proportions of children passing each of the 12 items, which was presumed to reflect item difficulty, did not differ significantly for either test. The possibility of differential response to male and female items on each test was investigated by subjecting the children's scores for male-target vs female-target items to Wilcoxon matched-pairs signed-ranks tests. No sex-of-target differences were found.

Test-Retest Stability. Stability of scores over a period of two weeks or less was assessed in two ways. Wilcoxon matched-pairs signed-ranks tests computed on the raw scores of the 20 children retested for these purposes revealed no differences for either test. In addition, the children were classified as having passed or failed each administration; one child who had passed the first attempt at the child test failed the second, and two children who had passed the adult test failed the second attempt. Phi coefficients computed to measure strength of association between first and second attempts at each test were $.87$ for the child test and $.83$ for the adult test.

Performance Effects

Sex Differences. Among subjects attempting the pretest, the proportion of boys who could not be tested was greater than the proportion of

untestable girls (32% vs 17%; $z = 1.98, p < .05$). None of these children were over 26 months of age. Age ranges and means for untestable boys did not differ from those for untestable girls.

The proportions of boys and girls passing either gender test at first attempt did not differ significantly; 28% of the boys and 30% of the girls passed the child test; 78% of the boys and 75% of the girls passed the adult test. To investigate the possibility of sex differences in responding at different portions of the overall age range, subjects were divided at the median age for each test (25.7 months, child test; 25.8 months, adult test); no differences in the proportions of younger or older boys and girls passing either test were found. Boys' and girls' scores were combined for the remaining analyses.

Age Differences. As expected, older children performed more successfully than younger ones. Fewer than 8% of the children below the median age passed the child test, while 50% above the median age passed ($z = 4.20, p < .0001$). For the adult test, 55% of the children below the median age failed but over 97% above the median age succeeded ($z = 4.47, p < .0001$). Proportions of children across the age range who passed each test are shown in Figure 1.

Test Differences. Because age differences in performance indicated a probable sequence for mastery of the two gender tests, the children who received both tests within one month ($n = 73$) were categorized as having passed both tests, failed both, or passed one but not the other. Of the 36 children who passed only one of the tests, just one child failed the adult test while passing the child test. McNemar's test for the significance of changes yielded a $\chi^2(1)$ of 30.25, ($p < .001$), confirming that the direction of change to be expected will proceed from mastery of the adult test to subsequent mastery of the child test. The same mastery sequence was found in informal retesting of 29 children who failed the child test at first attempt: 18 failed one or more subsequent attempts as late as 9 months after passing the adult

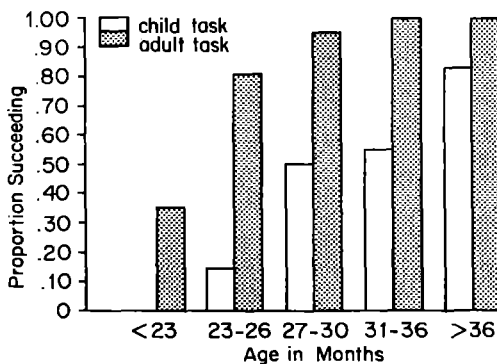


Fig. 1. Proportion of children across the age range succeeding on first attempt at each gender task.

test. Of six children retested on both tests, no child who passed the child test failed the adult test.

DISCUSSION

Children in this study demonstrated knowledge of gender labels for boys and girls as early as 24 months, and for adults even earlier. Each child who received the gender tests had first shown the ability to complete a similar discrimination task, ensuring that gender test failures were not caused by the nature or response mode of the task. Recognition rather than production of the labels is required, minimizing demand on verbal skills. This method of testing allows children's knowledge of labels to be assessed during the period in which language is developing and permits individual children to be classified as to their ability to discriminate men and women, and boys and girls, in response to the labels.

The gender tests are psychometrically adequate, easy to administer, and can be scored reliably. They could be extended readily to include other gender labels such as those found in Thompson's (1975) study, and used to investigate correlates and antecedents of label acquisition such as language proficiency, familial emphasis on sex typing, and the relationship between accurate labeling and behavioral sex typing. It would be desirable to obtain alternate sets of pictures for use when repeated testing is required and to enable children of other cultures to be tested with appropriate stimuli. The stimulus materials should be updated periodically to keep clothing and hairstyles reasonably current. Although the pictures used here were chosen deliberately to exclude toys or other objects that would provide gender clues, sex role knowledge could be tested by including such information or pitting it against the sex of the person portrayed.

As expected, older children generally performed better than younger ones on both tests and performance of boys and girls did not differ, but attrition among subjects under 26 months of age requires some caution in accepting the data for the very young children as normative. Those who could not or would not be tested may have been less likely to succeed, so that the data obtained may slightly overestimate the ability of the youngest children and, perhaps, fail to detect a sex difference. As more boys than girls simply declined to play the game, became upset when taken into the testing room, or could not or would not attend to the tasks, a greater proportion of girls than boys may have been willing to attempt the tasks before they had the ability to succeed. However, among children over 26 months of age boys and girls were equally willing to be tested and their scores did not differ.

The children who took part in this study clearly had a sex-based category system for adults before they could show discrimination of children. Many

identified pictures of adult males and females before the age of two, but could not yet identify boys and girls. Earlier success on the adult task is not surprising in view of the importance of parents to the child—these children could surely recognize their own parents, and label them as well (labels for parents are usually in use by 16 months of age; Brooks-Gunn & Lewis, 1979). In addition, adults are undoubtedly more discriminable than children in that they show greater differences in facial features, clothing, and hairstyle. But we know of no evidence that discrimination of adults is related to the young child's self-knowledge. Because identifying and labeling boys and girls appears a later and more variable accomplishment, tests of gender labeling or identity based on the discrimination of males and females that report data for adult and child stimuli combined (Thompson, 1975; Weinraub et al., 1984) cannot be interpreted as showing gender identity or overall mastery of gender labels.

The variability in age of mastery for the boy-girl task has no ready explanation. The social and intellectual development of these children was not in question. All were acquainted with peers of both sexes; indeed, the oldest children who failed have siblings of both sexes. The source of this variability may lie in the child's family experience, with parents—especially fathers—who are strongly sex-typed fostering earlier awareness of gender in their children, as Weinraub et al. (1984) has suggested. This and other potential correlates of gender labeling, such as verbal skill, invite further investigation.

Successful labeling implies nothing about the child's knowledge of the physiological basis of category membership, or of sex as a stable and permanent attribute, but it is important to note that success on either test implies categorical knowledge and possession of labels for the categories. At this point, children have begun to represent the categories symbolically and to be aware of what they know. Any female adult may be a mommy, but not a daddy, and vice versa. Children who can classify adults in this way may not be aware that the underlying categories are male and female, but they are classifying on the basis of gender nonetheless. However, children who recognize that diverse sets of boys and girls *are* boys and girls are dealing with categories to which they themselves belong. As gender knowledge appears mastered for self before other children (Eaton & Von Bargen, 1981; Gouze & Nadelman, 1980; Marcus & Overton, 1978), children who can label boys and girls accurately may be aware of their own sex as well. At the very least, those who can identify other children verbally as boys and girls are showing conscious awareness of the gender categories most relevant to themselves, and they could be expected to use this knowledge as they try to make sense of and fit into a sex-typed world.

Construction of these gender-labeling tests was motivated by the belief that the onset of accurate labels for boys and girls should be related to adoption of same-sex behaviors and rejection of opposite-sex behaviors in those areas over which the child can exert deliberate control, and was undertaken specifically in order to investigate this hypothesis. Fagot, Leinbach, and Hagan (1986) obtained behavioral observations of 43 children who took part in the present study. Three categories of behavior in which early sex differences have been found consistently—preference for same-sex playmates, aggression, and playing with sex-typed toys (Maccoby & Jacklin, 1974)—were observed in unstructured play and the behavior of children who passed the child gender test was compared with that of those who failed. Toy choice was not related to success at gender labeling; these children showed clear preference for own-sex toys regardless of their ability to identify boys and girls. Although preference for same-sex playmates was shown by both groups, children of both sexes who succeeded at the boy-girl task spent more time playing with peers of the same sex than did those who failed. The relationship between aggression and labeling was more complex. Girls who succeeded in identifying boys and girls displayed less aggression in the play group than girls who failed, but aggression scores of boys who passed the test, boys who failed, and girls who failed did not differ significantly. These children, most of whom were under three years of age, had already begun to make choices that brought their actions into accord with behavior expected of members of their sex, and for two of the three behaviors investigated, this was related to accurate responding to labels for boys and girls.

Two- to three-year-olds do appear to have more gender knowledge than they have been given credit for; although their gender concepts are rudimentary, they have begun to categorize people along gender-dimorphic lines. Moreover, the ability to categorize boys and girls appears related to behavioral choices. We believe gender labeling signals the point at which children are aware of discriminations they may already be making tacitly. If so, the power of gender-category names to permit inferences about category members, which Gelman et al. (1986) found in four-year-olds, may be present in even younger children and may serve to organize the child's gender knowledge or direct attention to sex typing in the environment. Gender labeling is indeed an important accomplishment, and as results of the present study suggest, there is a need for normative data on its attainment, for longitudinal testing to chart its course, and for further examination of its antecedents and correlates. The tests developed for this study meet the need for a simple instrument with which to assess gender labeling in children during the time of acquisition, and can contribute effectively to our understanding of sex role development in the very young child.

REFERENCES

- Blakemore, J. E. O., LaRue, A. A., & Olejnik, A. B. Sex-appropriate toy preference and the ability to conceptualize toys as sex-role related. *Developmental Psychology*, 1979, *15*, 339-340.
- Brooks-Gunn, J., & Lewis, M. Why Mama and Papa: The development of social labels. *Child Development*, 1979, *50*, 1203-1206.
- Constantinople, A. Sex-role acquisition: In search of the elephant. *Sex Roles*, 1979, *2*, 121-133.
- Eaton, W. O., & Von Bargen, D. Asynchronous development of gender understanding in children. *Child Development*, 1981, *52*, 1020-1027.
- Emmerich, W., Goldman, K. S., Kirsh, B., & Sharabany, R. Evidence for a transitional phase in the development of gender constancy. *Child Development*, 1977, *48*, 930-936.
- Fagot, B. I. Sex differences in toddlers' behavior and parental reaction. *Developmental Psychology*, 1974, *10*, 459-465.
- Fagot, B. I. The influence of sex of child on parental reactions to toddler children. *Child Development*, 1978, *49*, 459-465.
- Fagot, B. I., Leinbach, M. D., & Hagan, R. Gender labeling and the adoption of sex-typed behaviors. *Developmental Psychology*, 1986, *22*, 440-443.
- Fein, G., Johnson, D., Kosson, N., Stork, L., & Wasserman, L. Sex stereotypes and preferences in the toy choices of 20-month-old boys and girls. *Developmental Psychology*, 1975, *4*, 527-528.
- Gelman, S. A., & Markman, E. M. *Natural kind terms and children's ability to draw inferences*. Paper presented at the annual meeting of the Western Psychological Association, San Francisco, Calif., April 1983.
- Gelman, S. A., Collman, P., & Maccoby, E. E. Inferring properties from categories versus inferring categories from properties: The case of gender. *Child Development*, 1986, *57*, 396-404.
- Gettys, L. D., & Cann, A. Children's perceptions of occupational sex stereotypes. *Sex Roles*, 1981, *7*, 301-308.
- Gouze, K. R., & Nadelman, L. Constancy of gender identity for self and others in children between the ages of three and seven. *Child Development*, 1980, *51*, 275-278.
- Huston, A. D. Sex-typing. In P. H. Mussen (Ed.), *Handbook of child psychology: Vol. 4. Socialization, personality, and social development*. New York: Wiley, 1983.
- Kohlberg, L. A. A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences*. Stanford, Calif.: Stanford University Press, 1966.
- Kuhn, D., Nash, S. C., & Bruckner, L. Sex role concepts of two- and three-year-olds. *Child Development*, 1978, *49*, 445-451.
- Lewis, M., & Weinraub, M. Origins of early sex-role development. *Sex Roles*, 1979, *5*, 135-153.
- Maccoby, E. E. *Social development*. New York: Harcourt Brace Jovanovich, 1980.
- Maccoby, E. E., & Jacklin, C. N. *The psychology of sex differences*. Stanford, Calif.: Stanford University Press, 1974.
- Marcus, D. E., & Overton, W. F. The development of cognitive gender constancy and sex role preferences. *Child Development*, 1978, *49*, 434-444.
- Nelson, K. Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development*, *38* (Serial No. 149).
- Slaby, R. G., & Frey, K. S. Development of gender constancy and selective attention to same-sex models. *Child Development*, 1975, *46*, 849-856.
- Thompson, S. K. Gender labels and early sex role development. *Child Development*, 1975, *46*, 339-347.
- Weinraub, M., Clemens, L. P., Sockloff, A., Ethridge, T., Gracely, E., & Meyers, B. 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*, 1984, *55*, 1493-1503.