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

THE RELATIONSHIP BETWEEN UNDERSTANDING THAT GENDER IS UNCHANGEABLE AND THE DEVELOPMENT OF SEX-TYPED PREFERENCES IN PRESCHOOL-AGED CHILDREN

Brenda L. Kenyon

A Thesis in the Department of Psychology

Presented in partial fulfilment of the requirements
for the degree of Doctor of Philosophy at
Concordia University
Montreal, Quebec, Canada
January, 1994

ISBN 0-315-90857-2

Abstract

The Relationship Between Understanding that Gender is Unchangeable and the Development of Sex-Typed Preferences in Preschool-Aged Children

Brenda L. Kenyon, Ph.D.,
Concordia University, 1993

Gender constancy, the understanding that gender does not change when appearances do, has been proposed as an instrumental factor in the development of sex-typing. However, research has indicated that sex-typed preferences emerge by about four years of age whereas gender constancy is not demonstrated until about six years of age. Interestingly, by four years of age children generally pass appearance-reality distinction tasks (A-R Distinction), which assess the understanding that the identity of an object does not change when appearance does. In the present study, a test of gender constancy modeled upon A-R Distinction procedures was developed. The Appearance-Reality Gender Task (A-R Gender) consists of a series of photographs depicting a child dressing up like the opposite sex. It was hypothesized that by four years of age, children would demonstrate gender constancy by passing the A-R Gender Task, and that children who passed the A-R Gender

Task would have higher levels of sex-typing. Eighty-seven preschool aged children (28 three-year-olds, 34 four-year-olds, and 25 five-year-olds) each received the A-R Gender Task, the Sex Role Learning Index (SERLI) (Edelbrock & Sugawara, 1978), the Gender Constancy Interview (Slaby & Frey, 1975), and an A-R Distinction task. Contrary to predictions, the A-R Gender Task was not passed by the majority of children until five years of age. More children at each age level passed the standard A-R Distinction task than the A-R Gender Task, suggesting that gender posed more difficulty for children solving the appearance-reality distinction than did non-social objects. Interestingly, realism errors were most frequently responsible for A-R Gender failure after three years of age. Realism errors occur when children report the real gender of the pictured child rather than the apparent identity, indicating the understanding that gender has not changed. Children who passed the A-R Gender Task did not have higher levels of sex-typing, suggesting that cognitive competency with the concept of gender does not precede sex-typing. The implications of these findings for the role and measurement of gender constancy in the development of sex-typing, are discussed.

Acknowledgements

I would like to acknowledge the contribution of my supervisor, Dr. Lisa Serbin, who provided support and encouragement during the course of this research and my studies at Concordia. I am indebted to Dr. Diane Poulin-Dubois for her enthusiastic involvement in this project from its inception to completion. I am grateful to the members of my examining committee, Drs. Norm Segalowitz, Lynn Zarbatany, and Patricia Lightbown, for their comments and contributions to the final product, and for having made the oral defense an exciting and rewarding experience. I would also like to extend my thanks to Jacky Boivin, who provided thoughtful consultation during the process of data analysis, and to Lucy Brunetti for her assistance with data collection.

I would like to thank the staff and students of Loyola Daycare Centre, Garderie Notre Dame du Grace, Garderie Mont Royal, St. Andrews School, Child's Preparatory Day Care Centre, Snowdon YWCA Day Care Centre, and University Heights Day Care, for their participation in this study.

A group of extraordinary people gave meaning to my years in Montreal, and to them I express my heartfelt gratitude. I shared many joys and sorrows with Susan Graham, Patricia Peters, Steven Lapidus, and Betsy and Kevin Austin. Not only do I love them deeply, but they are passionate, creative people for whom I have a great deal of

respect. I am honoured to be their colleague, treasure their friendship, and look forward its continuance in the years to come.

I would also like to acknowledge the contribution of those people who provided tremendous support during the difficult year prior to the completion of this project. Linda Kenyon and Bob Quick continue to be valued and loved life companions. Linda also gave generously of her editorial skills, which I (and my readers) appreciated. Heather Rennie thoughtfully read drafts of this document, as well as provided ongoing personal support; I appreciate her as a person of spirit and strength. Charles Evans provided help and encouragement during the last year and, during the preparations for oral defense, was as supportive and understanding a partner as one could want. I look forward to returning the favour with the same generosity. Finally, I would like to acknowledge the continued support of my dear friend Bonnie Lewis who, for years, has encouraged me to do what I love and to love what I do. And I do.

Dedication

This thesis will forever remind me of the years in Montreal and the difficult lesson I learned about being loved, about my own strength, and about courage. This I dedicate to my husband Charles Evans, who is becoming my model of courage, as we look to our future together.

Table of Contents

List of Tables

Introduction

In every human culture, gender serves as a basic organizing principle for information about activities and interests, personal-social attributes, social roles, and relationships (Bem, 1981). The process by which children internalize gender information and acquire behaviours, attributes, roles, and values associated with being male or female is called sex-typing. Sex-typing is an important feature of children's personal and social development and children's gender concepts have received extensive attention in the pursuit of understanding how sex-typing develops.

In her comprehensive review of research in the development of sex roles, Huston (1983) listed five major components identified with the concept of gender: 1) biological sex, or, the distinction between male and female; 2) interests and activities; 3) personal-social attributes or characteristics; 4) gender-based social relationships, roles, and behaviours; and, 5) stylistic and symbolic characteristics associated with masculine or feminine. Although the specific content and relative importance of each of the components may vary between individuals or cultures, they constitute much of our concept of gender. Gender concepts begin to form early in infancy and almost all the of the components Huston identified are mastered in the preschool years.

The early understanding of the first component, the

basic distinction between male and female, has been demonstrated in the first year of life when, using a preference-looking procedure, 12-month-old infants have been shown to match male and female voices with male and female faces (Poulin-Dubois, Serbin, Kenyon, & Derbyshire, in press). The understanding of gender labels, such as lady, man, girl, and boy, has been demonstrated as early as 26 months of age, when children have been found to reliably produce the correct labels for pictures of people (Weinraub, Clemens, Sockloff, Ethridge, Gracely, & Meyers, 1984). By 31 months of age, children can reliably point to pictures that match gender labels (Fagot & Leinbach, 1985; Etaugh, Grinell, & Etaugh, 1989). These data indicate that by the end of the second year infants can clearly discriminate the two sexes, and can identify at least three of the features (voices, faces, and social labels) associated with each.

The second component identified by Huston (1983), activities and interests, is also evident in early infancy. In a study of 10-month-olds, Levy, Haaf, and Sommer (1991) found that infants habituated to pictures of males and females matched with gender-stereotyped objects and then generalized habituation to pairings of the objects with other appropriate-sex faces, suggesting an association between gender and social information. Similar results have been found with older children. Using a sorting procedure, 24-month-old children have demonstrated knowledge of the

stereotypes of toys and activities associated with boys and girls (Kuhn, Nash, & Brucken, 1978; Warren & Von Bargen, 1981; Weinraub et al., 1984; Martin, 1989).

Knowledge of the third component, personal-social attributes, is also evident in early childhood. Kuhn, Nash, and Brucken (1978) found that 30-month-olds attribute personal characteristics (e.g., "says 'I need help'" or "says 'I can do it'") in a sex-typed fashion, indicating the early expectations of sex differences in personal-social attributes. The fourth component, understanding gender-based relationships, has been demonstrated at three years of age when children have been found to react differently to male and female adults in free-play situations (Fagot, 1984) and to make gender-based playmate choices (Serbin & Sprafkin, 1986).

Investigations of children's understanding of the final component, gender-based stylistic and symbolic features, are less prevalent. However, there is some evidence that Grade 1 children are sensitive to the culturally-shared symbolic associations with masculine (e.g., fast, loud, hard) and feminine (e.g., slow, quiet, soft) in such forms as television commercials (Welch, Huston-Stein, Wright, & Plehal, 1979).

Taken together, these studies indicate that relatively sophisticated gender concepts develop early in the preschool years. Curiously, despite their sophisticated gender

concepts, preschool aged children do not appear to understand that gender is a permanent attribute. Before about six years of age many children believe that gender can change. For example, a preschool-aged girl might say she is a girl but was a boy yesterday, or that she would become a boy if she played with boys' toys. It is not until about six years of age that most children correctly report that their gender will not change when such features as appearance or activities do. The understanding that gender does not change has been proposed to be an important feature of children's gender concepts, one that plays a critical role in the development of sex-typing.

Cognitive-Developmental Model of Sex-Typing

Kohlberg (1966) proposed a cognitive-developmental model of sex-typing in which changes in children's concept of gender were considered instrumental in sex role development. Kohlberg suggested that the categorization of oneself as male or female and acquiring an "unchangeable sexual identity" determined the subsequent valuation of masculine or feminine behaviour as desirable and appropriate for oneself (1966, p. 83). A critical feature of Kohlberg's proposed model is the development of the understanding that gender does not change, which is called gender constancy.

Kohlberg (1966) proposed that gender constancy develops in three stages. First, children understand that people are

either male or female. Next, they learn that boys become men and girls become women. Finally, children appreciate that gender does not change even if physical or behavioral features do. Slaby and Frey (1975) developed a measure of gender constancy, the Gender Constancy Interview, which consists of questions reflecting the stages of gender concept development proposed by Kohlberg: Children are first asked questions of gender identity ("are you a boy or a girl?"), then stability ("when you were little, were you a boy baby or a girl baby?" / "when you grow up will you be a mommy or a daddy?"), and finally, constancy ("if you wore [opposite sex] clothing would you be a boy or a girl?"). For each question, children are asked a counter question intended to assess the stability of their response. For example, after the question "are you a boy or a girl?", they are asked "Are you a [opposite sex of child]?", and would have to correctly deny it to pass the identity item. The Gender Constancy Interview or some variation of it (i.e. Emmerich, Goldman, Kirsch, & Sharabany, 1977) has been used extensively to assess children's gender concepts.

Research has indicated that the ability to pass gender constancy tasks is a robust developmental phenomenon and emerges in Kohlberg's predicted stages. Using the Slaby & Frey Gender Constancy Measure, research has demonstrated that children pass Identity items by about three years of age (Fagot, 1985; Taylor & Carter, 1987). Stability items

are usually passed by four years of age (Fagot, 1985; Taylor & Carter, 1987) although they have been passed by children as young as three and one-half years (Coker, 1984). Constancy items are not usually passed until about six years of age (Coker, 1984; Taylor & Carter, 1987) although they have been passed by children as young as four years of age (Slaby & Frey, 1975). Performance on gender constancy tasks improves at a surprisingly regular and consistent rate, and Guttman Scalogram analyses have indicated that the sequence is relatively invariant (Kuhn, Nash & Brucken, 1978; Coker, 1984). Thus, children appear to develop gender constancy in the manner predicted by Kohlberg.

Gender Constancy and the Development of Sex-Typing

The understanding that gender does not change features prominently in Kohlberg's (1966) model of the development of sex-typing. He proposed that the process begins with the acquisition of gender information, which begins early in infancy as soon as children understand that there are two sexes. Children then develop the understanding that they are one sex or the other. This understanding, called gender identity, is based upon the physical reality judgement of whether the child is male or female, and is first understood basically as a nominal tag with no relevance for children in terms of behaviour. Thus, at about two and a half years of age, children label themselves male or female much as they

label themselves by their name but are not motivated to use their gender knowledge to guide their own behaviour.

According to Kohlberg (1966), the next influence on the development of sex-typing is that of the crystallization of children's gender concept to include the understanding that gender is a stable, unchangeable attribute. Only after children acquired an unchangeable sexual identity were they believed to feel motivated to adhere to their perceptions of what is sex-appropriate. The process by which constancy was proposed to affect sex role development was thought to emerge from the need to preserve a positive sense of self. Kohlberg, citing White's (1959) model of identity formation, believed that children are motivated to preserve a sense of general competence, effectiveness, and self-regard and accomplish this by positively valuing things associated with themselves. Thus, it was proposed that when a child identified his or her gender as an immutable characteristic, it would become a valued, self-defining feature. Kohlberg believed that a child's sexual identity, once crystallized, was developed and expanded upon by the "motivated adaptation to physical and social reality" (Kohlberg 1966, p. 88). Children learn sex role information from their active processing of information in the environment; however, the stable self-identification of being male or female motivates children to actively attend to and value sex-appropriate information. Differential modelling studies have offered

some support for this process, demonstrating that only after children have acquired gender constancy do they reliably attend to and imitate same-sex models (see for example, Bussey & Bandura, 1984).

Although Kohlberg's (1966) proposal has generated a great deal of interest among researchers, the literature to date offers little support for the role of gender constancy in the development of sex-typed preferences. No relationship has been found between levels of gender constancy and sex-typed preferences for toys (Coker, 1984; Carter & Levy, 1985, 1987; Downs & Langlois, 1988; Martin & Little, 1990), for games or activities (Marcus & Overton, 1978; Downs & Langlois, 1988), or for same-sex playmates (Marcus & Overton, 1978; Serbin & Sprafkin, 1986; Martin & Little, 1990). In fact, sex-typed preferences are generally evident by four years of age, two years before children can be expected to pass gender constancy measures (Weinraub et. al., 1984; Emmerich & Sheppard, 1984).

The failure to find a relationship between gender constancy and sex-typed preferences can be explained in three ways. First, understanding that gender does not change may, indeed, be unrelated to the acquisition of sex-typing. However, it is difficult to reconcile the attention and strong affectional ties that preschool children devote to sex-appropriateness with their belief that sex is transient and could change at any time.

A second explanation emerges from studies of the Slaby and Frey (1975) Gender Constancy Interview, which indicate that the measure is problematic. Many children may fail the Gender Constancy Interview because of the counter questions used to assess the stability of the subject's responses. Young children may be influenced by the counter questions, experiencing them as criticism or doubt on the part of the examiner, and may change their answers. Indeed, the use of counter questions in non-testing situations most probably occurs only when a child's statement is incorrect or doubted (e.g. Adult: "Did you pick up all your toys off the floor?" Child: "Yes". Adult: "Did you leave any on the floor?"). Particularly for young children, counter questions may induce them to change their responses. In a study of preschool children's perception of the Gender Constancy Interview, Seigal and Robinson (1987) had children listen to the story of a fictitious child passing the Constancy questions but failing the counter questions. When asked why they thought the fictitious child had changed his or her answer, children frequently responded that they thought the examiner had indicated that the first answer was wrong.

The type of questions posed in the Gender Constancy Interview may also be difficult for young children. MacKain (1989) found that hypothetical gender constancy questions, the type used in Slaby and Frey's interview, were more difficult for preschoolers to answer than were those with a

concrete referent. Thus, the Slaby and Frey instrument may underestimate children's understanding of gender.

The third, and most important possible reason that no relationship between gender constancy and sex-typing has been found concerns Kohlberg's conceptualization of the measurement of gender constancy as a conservation task. Kohlberg (1966) conceptualized gender constancy in terms of conservation, the conservation of gender across changes in perceptual characteristics. He proposed that gender constancy was "one aspect of the stabilization of constancies of physical objects that takes place between three and seven years" (Kohlberg, 1966, p. 90). Gender constancy was, therefore, thought to reflect growing cognitive competence with conservation of physical properties. However, Kohlberg (1966) failed to distinguish between quantitative and qualitative invariants when assuming the parallel development of constancies. The difference between qualitative and quantitative invariance is subtle but important and has implications for the assessment of children's understanding of transformations.

Qualitative vs. Quantitative Invariants

According to Piaget (1968), qualitative and quantitative invariants differ in the nature of invariant attributes and the cognitive functions required for invariance. Quantitative invariants are composed of

differences that directly compensate each other, for example, as the height and width of water in a tall thin or a short wide beaker compensate each other. In order to master quantitative invariants, children must understand the rules for performing equivalence operations (e.g., that water level will change proportionate to the width of the beaker). Qualitative invariants differ in that they are non-quantifiable (e.g., colour, identity) and have no direct compensatory equivalents. To understand qualitative invariance, children need only recognize the continued physical existence of objects, despite apparent changes in size or shape. The invariant identity can be understood without quantitative composition: children need only disassociate the permanent from the variable qualities. Piaget (1968, p.18) used conservation to refer to quantitative invariance only. Thus, object permanence, which is sometimes considered a precursor of conservation (e.g., Siegler, 1991), is actually a qualitative invariance task because the occluded object is still expected to "be."

Although pre-operational identity structures, such as object permanence, are sometimes considered the basis for conservation of quantity, the two are not the same. Qualitative invariance requires only the understanding that things continue to exist despite perceptual changes whereas quantitative invariance requires the understanding of how changes in features of objects compensate for changes in

other features. In his discussion of the difference between qualitative and quantitative invariants, Piaget (1968) described children's ability to separate the two forms of invariants within the same object. When young children were shown a wire that was then crunched up into a ball, they believed that it was the same wire (qualitative invariance) but that its length had changed (quantitative invariance), indicating that they knew that the same wire existed but could not perform the operations necessary to determine that the wire maintained the same length. Qualitative invariance emerges earlier and is separate from quantitative invariance, and the distinction between them is important to the measurement of children's understanding of identity transformations.

With few exceptions, the procedure used to test children's understanding of qualitative invariance of identity is derived from that used to test conservation of quantity. Children are usually shown an object, its appearance is changed, and they are then asked to report its true identity. Before the age of about six or seven years, children perform poorly on such tasks. For example, DeVries (1969) had preschoolers observe a very patient cat that had its face covered with a realistic-looking dog mask. Children were then asked what the animal really was. Not until five years of age did most children report that it was still a cat. Similarly, Keil (1979) presented preschool-

aged children with stories of a horse who was dressed in a costume and looked like a zebra. He found that before the age of five years, children believed that changes in the appearance changed the identity of the animal, although when the costume was emphasized, several children allowed that costume changes were not sufficient to change the horse into a zebra. The same procedure, of changing appearances and asking children to report identity, has been used to test children's understanding of the effect of name changes on identity (Guardo & Bohan, 1971), changes of clothing on ethnic identity (Aboud, 1984), and "transplants" of body parts on identity (Johnson, 1990). In each case, children under about six years of age report that changes in such features will result in changes of identity.

The procedure used in these studies involves three important variables. Using DeVries's (1969) example of the cat dressed as a dog, the variables could be described as follows. The first variable (A) is that of the true identity, the cat. Variable B is those features that are invariant and are the consequences of being a cat (e.g., has cat bones and blood, has kittens etc.). Variable C is the variant features (such as having whiskers or a long tail), which may be shared by some cats but not all. Among these three variables, the only imperative relationship is that between A and B: Given that this creature is a cat, it must have the invariant cat features. Variable C is unreliable

and may vary across types of cats. Knowledge of A (that the animal is a cat) does not automatically predict C (that it will have whiskers or a tail). However, it is precisely this association between identity and variant features which is tested using conservation procedures. The unreliable perceptual features are changed (C) and children are asked to infer the identity (A) of the "new" animal.

Consider the same procedure when it is appropriately used for assessing children's understanding of the compensatory nature of quantitative features. Tests of quantitative invariance (e.g., transferring a mass of sand from a tall thin beaker to a short wide one) have only two important variables: the object's implied identity (e.g., a single mass of sand) and its features (e.g., height in the beaker, weight, volume, etc.). There are no predicted consequences of knowing the identity of the object. Knowing that it is a mass of sand predicts nothing about the shape it will take, the area it could cover, or the weight it might have. One would need further information about the specific features of this mass of sand, such as its volume or weight, which would not necessarily be true of other masses of sand. Knowing that it is a mass of sand predicts no other information except, perhaps, its own qualitative invariance: no matter how it is changed perceptually, it will still be a mass of sand. Asking children to decipher "sameness" when presented with changes to the variant,

changeable features, accurately tests their understanding of equivalences, or compensatory nature, as described above. However, it is a weak test of qualitative invariance.

Using quantitative invariance procedures for assessing children's understanding of the unchangeability of identity ignores the essential defining features of the identity in question and instead requires children to generate an identity from the collection of variable, unreliable, and nonessential properties.

The gender constancy interview derived from Kohlberg's proposal involves exactly this procedure. Children presented with changes in a child's appearance and asked what sex the child is are essentially being asked to generate identity based upon changes in nonessential, unreliable features. The understanding that identities do not change with changes in perceptual information is better tested by presenting children with the identifying information (A in cat example), changing the nonessential features (C in cat example) and asking children if, given the object's identity, the essential features remain the same (B in cat example). This procedure has been used to study children's use of category labels, and the results are strikingly different from those obtained with standard conservation procedures.

In a series of experiments, Gelman and her colleagues investigated children's ability to use category labels to

organize information about objects. The procedure used, called a property-inference task, involved providing the category label for an object that was perceptually misleading and asking children to make a decision about what essential features the object would have, that of its apparent or given identity. In one study children learned a new fact about both dinosaurs and rhinoceroses (that the first has cold blood while the second has warm blood) (Gelman & Markman, 1986). They were then shown a picture of a dinosaur that looked like a rhinoceros. Children were told that the animal was a dinosaur, and were asked if they thought it had warm or cold blood. Sixty-eight percent of four-year-old children responded correctly, indicating that they used the animal's category, rather than its appearance, to guide their inferences. In a similar study Gelman and Coley (1990) presented pictures of a typical bird (bluebird), an atypical bird (dodo bird), and a dinosaur which resembled a bird (pterodactyl) to two-year-old children. The children then learned facts about birds (e.g., lives in a nest, lays eggs). The atypical bird and perceptually-similar dinosaur were then presented and named for the children, who were asked whether it would or would not have the bird-like features. Gelman and Coley found that as early as two and a half years of age, children made correct inferences about features of the atypical category members, indicating that they could efficiently overlook the

conflicting perceptual information. These studies provide evidence to suggest that children understand that identity is not perceptually-bound, and similar results have been obtained with the concept of gender identity.

Gelman, Collman, and Maccoby (1986) had four-year-old children learn new facts about boys and girls (e.g., that boys have andro in their blood and girls have eggs inside their bodies). Children were then presented with a picture of an ambiguous-sex child and told it was either a boy or a girl, then asked whether it had andro in its blood or eggs inside its body. Gelman and her colleagues found that children willingly inferred non-obvious properties about the ambiguous child when provided with its gender label. This indicates that, for young children, gender is a category which is expected to predict important features and is not affected by conflicting appearances.

These studies clearly indicate that when they are told an object's identity, children as young as three years of age ignore conflicting perceptual information to make inferences about essential features. Given these data, and the problems associated with the Gender Constancy Interview, the apparent lack of gender constancy evident in young children may be due to methodological problems in the measurement of the understanding that gender does not change. Thus, it remains possible that preschool aged children do not believe that gender will change when

features such as appearance or behaviour are changed.

It could be argued that the two procedures described above, that of asking conservation-type transformation questions and that of asking subjects to make inferences about properties given the categories, both elicit specific performance from young children. For example, asking children to determine what an animal is after having witnessed changes to nonessential features may suggest to children that, indeed, identity could have changed. In this respect the question is anomalous and may prompt children to use perceptual features to generate a new identity through a form of "bottom-up" processing. Alternatively, it could be argued that property-inference tasks encourage children to ignore the conflicting perceptual information by drawing attention away from the appearance of an object and directing it to the label. What remains to be explained is what preschool children believe to be the nature and the consequences of changes in nonessential features. Recent research in the developing cognitive capacities of preschool aged children has provided information about children's understanding of such transformations, as well as providing a potential methodology for re-assessing the question of when children understand that gender does not change. The procedure is that used in the study of the appearance-reality distinction.

Appearance-Reality Distinction

Theory

The appearance-reality distinction, which has been studied extensively by Flavell and his colleagues, refers to the understanding that objects can be represented (experienced, perceived) in different ways. Flavell, Green, and Flavell (1986) proposed that the ability to distinguish appearance from reality reflects a change in children's cognitive capacities regarding perception of the world. Flavell described this change as a transition from Level 1 to Level 2 knowledge about visual perception. Level 1 knowledge involves the understanding that people are cognitively connected to objects and events in the world. By two or three years of age, children know that objects can be seen, tasted, or felt. They are "sophisticated seers" who can show things, hide things, and know when other people can or cannot see something. As well as knowing that objects can be perceived, very young children know that objects can be mentally represented --thought about, remembered, wanted-- when not in view. By three years of age, children's understanding includes knowing that mental representations are not real, cannot be touched or seen, and are not consistent over time (Wellman & Estes, 1987). However, according to Flavell (1988), children at this level lack the important understanding that mental representations can differ between people. They characterize objects as

having only one nature at any given point in time and cannot understand how something could be perceived, experienced, or described differently from their own perception of "the way it is." Understanding that objects can be mentally represented differently from one's own experience is called Level 2 knowledge. Level 2 knowledge enables children to understand that an object can be mentally represented both in different ways by the same person or differently by two different people.

According to Flavell (1983), the transition from Level 1 to Level 2 thought is a monumental change in children's development of metacognition, which is the ability to think about their own and others' thoughts. It represents the move from knowledge based upon cognitive connections to knowledge based upon mental representations. The importance of understanding that other people could have different mental representations is clear given that a substantial amount of competence as social beings depends upon this ability. With the development of Level 2 thinking children learn that other people truly can "see things differently," can have mistaken beliefs (Seigal & Beattie, 1990), or can be motivated to act upon previous knowledge or beliefs (Moses & Flavell, 1990). The development of the understanding that mental representation of objects can differ from their real identity is an extremely important skill, and is illustrated in children's performance on

appearance-reality distinction tasks.

Appearance-Reality Distinction Research

In appearance-reality distinction tasks, children are shown an object that looks like one thing (e.g., a rock) but really is something else (e.g., a sponge that has been painted to look like a rock). Children are first shown the object and asked what it looks like. Then they are given the object and its true identity is revealed. The object is again held in front of them and they are asked what it "really is." Studies of appearance-reality distinction have indicated that competence on appearance-reality distinction tasks emerges between the ages of three and five years.

Flavell, Flavell, and Green (1983) asked children to answer appearance-reality questions about four objects (a sponge that looked like a rock, a piece of rubber that looked like a pencil, a Charlie Brown puppet that was then covered with a sheet to look like a ghost, and a white card that was held under a pink filter to look pink). Consistent with the expectations associated with Level 1 and Level 2 thought, three-year-old children did not pass the appearance-reality task at above chance levels. Only 25% of the three-year-olds passed, whereas 85% of the four-year-olds and 100% of the five-year-olds passed.

When solving appearance-reality distinction tasks, children can err in three ways. They can make appearance-

based answers to reality questions (phenomenism errors) by saying, for example, that the sponge looks like a rock and really and truly is a rock. They can also make reality-based errors to appearance questions (realism errors) by saying, for example, that the sponge really and truly is a sponge, and looks like a sponge. Finally, children can respond randomly to questions in which case they can make both errors on any one object.

Interestingly, Flavell and his colleagues found that children who failed the appearance-reality tasks did not err randomly. Few children failed both questions associated with any one object. Furthermore, children did not consistently make phenomenism errors, indicating that they were not routinely misled by appearances. Instead, children either correctly reported the appearance then persevered with that answer, or correctly identified what the object really was and persevered with that answer. Overall, children made equal numbers of phenomenism and realism errors. When objects were examined separately, however, it was found that phenomenism errors were made most frequently on the task involving the colour transformation of the white card. Flavell et al. explained the difference in errors in terms of different task demands associated with the white card, because it differed from the rest in the nature of the transformation tested. Colour transformations are property transformations, in which a non-essential aspect of the

object is changed. The white card covered with a pink filter still looks like and really is a card. However, the queries are regarding a feature of the card (its colour) not its identity. The appearance changes in the other tasks, however, involved changes in the apparent identity of the objects. In a subsequent study, Flavell et al. (1983) tested both property and identity changes of the same objects, with apparent size changes (using a microscope), colour changes (using filters), and identity changes of several objects. Changes of identity yielded more realism errors than did changes of properties, which elicited more phenomenism errors.

Flavell and his colleagues (1983) suggested that these differences occurred because children's performance could be characterized as identifying the 'true state of affairs,' using whatever feature they identified with the object, and persevering with that response. For example, in a property task where the object was presented and identified by its colour, colour may have become the essential identifying feature. Therefore, when its colour changed, children may have assumed that its identity had. With identity tasks, however, the identity of objects may have been linked to features other than appearance (e.g., the sponge's texture or the pencil's uses) which did not change. When asked the appearance-reality questions, children, knowing that appearance is not sufficient to predict identity, may

persevere with their perception of the 'true identity' resulting in different errors for identity and property items. In each case, children may have determined identity of the object based upon salient features that differed for property and identity tasks and persevered with that response. It would be interesting to have this hypothesis tested using an identity task that involves only a change of colour, for example, changing a lemon into a lime, or cherries into grapes using filters.

Summary of A-R Distinction Studies

At three years of age, the ability to distinguish between appearance and reality is "precarious" (Flavell et. al., 1983), and neither specific training nor feedback appears to improve it. Although some three-year-olds correctly solve appearance-reality distinction tasks, not until four years of age do most children reliably report both the apparent and real identity of objects. Flavell and his colleagues concluded that the data indicate that the belief that children accept perceptual characteristics as identity is inaccurate. The findings suggest that by four years of age, most children can successfully report both what an object is and what it looks like, indicating the ability to represent both aspects of a transformed object and to differentiate the perceptual identity from the real one.

The studies reviewed above render a complex picture of children's related competencies with identity transformations, which can be summarized as follows. Although children do not acquire the necessary understanding to perform conservation of quantity tasks until about six or seven years of age, qualitative invariance emerges much earlier. By three years of age, children use the known identity to guide inferences about objects even when appearance is misleading. This suggests that children do not believe that identity changes when appearance does. However, children routinely fail traditional transformation questions until about five or six years of age, perhaps because these tasks require the direct confrontation of phenomenal and real identities. Research with appearance-reality distinction tasks has indicated that when children are specifically queried about the difference between the real and apparent identities of transformed objects, they can competently report the differences by four years of age.

Given that gender is an important category for children, one with which they develop considerable competence at an early age, it is unlikely that children's failure on the Gender Constancy Interview at five years of age reflects the inability to appreciate that gender does not change when appearances do. It is more likely that failure reflects preschoolers' difficulty disassociating the real from the apparent identity when reporting what the

dressed-up child "really is." The appearance-reality distinction procedure therefore provides a method for assessing children's understanding of the apparent transformation of gender, one which could provide important information about both preschool children's gender concepts and why they consistently fail transformation questions.

Relationship Between A-R Distinction and Gender Constancy

The appearance-reality distinction procedure is an appropriate method for determining when children understand that gender does not change, for several reasons. First, the procedure is methodologically appropriate for young children. Unlike the Gender Constancy Interview counter questions, which may be difficult for three-year-olds, appearance-reality distinction tasks have been successfully used with preschool children. Second, like gender constancy questions, appearance-reality questions assess children's understanding of a transformation: the object looked like one thing but was changed to look like something else. Third, appearance-reality distinction questions are about objects which can be seen and with which children can have some physical contact, thus avoiding the difficulties associated with hypothetical questions described earlier. However, the Gender Constancy Interview (Slaby & Frey, 1975) uses questions that are hypothetical and abstract. The use of a concrete referent may facilitate young children's

understanding of the questions and the task. Finally, the appearance-reality distinction procedure is appropriate for examining children's understanding that gender does not change because previous research has provided a base of information about preschoolers' performance on appearance-reality distinction tasks, against which gender-related tasks can be compared. Thus, the distinction between performance on non-social objects, such as rocks and sponges, can be compared with performance on a socially-laden construct, such as gender. To date, only two investigations of appearance-reality distinction task performance and gender constancy have been undertaken, and these have only focused upon the relationship between the two abilities.

Studies of Gender Constancy and the Appearance-Reality Distinction

Trautner (1985) examined the performance of 130 school-aged children (ages three years, eight months to nine years, five months) on measures of gender constancy and appearance-reality distinction. Trautner's measure of gender constancy consisted of 21 questions assessing various components of the concept of gender (identity, stability, constancy); however, the questions varied between questions about children's own gender and questions about others' gender. Children had to provide explanations for their answers in

order to be credited with passing.

The appearance-reality distinction task consisted of four items: line magnification (with a magnifying glass), the colour change of a white card (using a coloured filter), the relative change in the length of a line (using optical illusion), and the disguising of a puppet by covering it with a white cloth. Children were classified as realists (passed all four tasks) nonrealists (failed all four tasks) and transitionals (passed up to three tasks). Children received the appearance-reality distinction task either with corrective feedback or without feedback. The nature of the feedback was not specified.

Trautner (1985) found that all children who passed gender constancy also passed appearance-reality distinction. Similarly, none of the children who failed all the appearance-reality distinction tasks passed gender constancy. Thirty-seven children passed appearance-reality distinction but not gender constancy. Trautner also reported that children who passed appearance-reality distinction tasks had the "same gender constancy scores, regardless of age," although the age distribution is not given.

From these results, Trautner (1985) concluded that the skills underlying appearance-reality distinction preceded the emergence of gender constancy and are therefore necessary but not sufficient skills for the acquisition of

gender constancy. However, this conclusion must be accepted cautiously because the performance of the children in this sample was not equivalent to that found in other studies with similar measures. From a frequency table provided it can be calculated that the average age of children passing gender constancy was seven years and three months old, comparatively older than has been found. This difference may have been due to the requirement that children explain their answers, which has been shown to be difficult for children under six or seven years of age, unrelated to their concept of gender (Emmerich et al., 1987). The finding that appearance-reality distinction was a necessary precursor of gender constancy may have been an artifact of the advanced age of the gender constant sample: considering that appearance-reality distinction is usually passed by about four years of age, it would be surprising if children over seven years had failed it. The children in this sample also passed appearance-reality distinction later than has been found elsewhere. Only 36% of the four-year-olds and 83% of the six and seven-year-olds passed. Why these children performed so poorly on this task is not clear.

This study would have been greatly improved by an analysis of children's performance on individual objects. Three of the four objects in this study were property-change tasks, which, as discussed before, may provide information about children's difficulty assessing the critical features

that constitute identity in lines and shapes. However, it is the ability to recognize that the identity is unchanged when appearances change that is critical to gender constancy. Therefore, identity tasks would have been a more relevant source of comparative information about these children's functioning. Unfortunately, which objects were passed and failed most frequently were not reported in this study. Despite these problems, Trautner's (1985) finding that none of the children who failed appearance-reality distinction passed gender constancy suggests that appearance-reality distinction is related to gender constancy. These findings stimulated a second study that discriminated between children's understanding of gender constancy, for the self and for others.

Brown and Pipp (1991) examined the relationship between gender constancy and appearance-reality distinction in a sample of 77 children ranging in age from two-and-a-half to six years old. Children received the Slaby and Frey (1975) measure of gender constancy and were not required to provide explanations for their answers. The gender constancy questions included a series of questions regarding constancy for self and others. The ability to distinguish appearance from reality was determined by performance on four tasks. Two were identity tasks (a sponge that looked like a rock, a wooden ball that looked like an orange), and two were property tasks (a line that looked crooked or straight, a

white card under a red filter). Children received one point for every correct response in either task, yielding an average gender constancy and appearance-reality distinction score.

Analysis of the children's scores on the measures indicated that appearance-reality distinction scores were lower than gender constancy scores, suggesting that gender constancy was passed before appearance-reality distinction. When the distinction between children's understanding of their own gender and that of others' was made, the results more closely resembled those of Trautner (1985). It was found that gender constancy for the self was passed before constancy for others and that self-constancy was not necessarily related to the ability to pass appearance-reality distinction tasks. Understanding constancy for others, however, was related to the ability to pass appearance-reality distinction tasks. Brown and Pipp concluded that understanding that one's gender does not change develops before the understanding that other people's gender does not change and suggest that cognitive abilities such as the appearance-reality distinction are necessary only for developing constancy for others.

Unfortunately, the average age of children passing each task is not reported in this study, and performance on the property and identity tasks was not analyzed separately. Therefore, it is not possible to determine whether or not

children's performance on the identity tasks was related to gender constancy. Nonetheless, the results suggest that appearance-reality distinction may be a related ability to the development of gender constancy.

Summary of A-R Distinction / Gender Constancy Studies

Taken together, the two investigations described above provide some support for the possibility that the skills required for understanding the appearance-reality distinction may underlie the ability to solve gender constancy transformation questions. However, before this can be confidently concluded, two important areas need be addressed. First, there is evidence that performance on property tasks is different from performance on identity tasks. Gender constancy is essentially an identity task. Thus, the relationship between gender constancy and performance on appearance-reality distinction **identity** tasks should be examined. When the understanding of identity transformations alone is tested, children's performance may be similar on both appearance-reality and gender constancy tasks. Second, given the measurement problems associated with the Gender Constancy Interview and the suitability of appearance-reality distinction tasks for testing children's understanding of transformations, it would be informative to test children's understanding that gender does not change when appearances do, using an appearance-reality distinction

procedure. Using this method of assessing children's understanding of the unchangeability of gender, Kohlberg's (1966) proposed model of the development of sex-typing can be re-addressed. This was the goal of the present study.

Present Study

The purpose of this study was to re-address the question of when children understand that gender does not change and to determine whether or not that understanding is related to the development of sex-typed preferences. It is important that the question of when children understand that gender is unchangeable be re-addressed, both for its relevance to understanding children's gender concepts and for its possible relationship to sex-typing. If the understanding that gender is unchangeable develops earlier than has been found, it may in fact coincide with the emergence of sex-typed preferences as Kohlberg proposed, thus providing evidence of the mechanism by which sex-typing emerges. In the literature the term 'gender constancy' is usually used to describe performance on the Slaby and Frey (1975) measure. Thus, 'gender constancy' has been used to refer to task performance which, for reasons discussed earlier, may not accurately reflect children's understanding of the unchangeability of gender. In the present study, 'gender constancy' is reserved for referring to performance on Slaby and Frey's (1975) Gender Constancy Interview.

Goals of the Study

There were two major goals in the present study. The first goal was to develop a method to assess the understanding that gender does not change. A task was developed (the A-R Gender Test) that involves a series of photographs depicting gender transformation. The transformations involved a boy dressing up as a girl and a girl dressing up as a boy. The pictures were accompanied by questions based upon appearance-reality distinction task procedure. Standard measures of gender constancy (Gender Constancy Interview, Slaby & Frey, 1975), and appearance-reality distinction (Flavell, et al., 1983) were also used to enable the comparison of performance on the A-R Gender Test with performance on other related tasks.

The second major goal of this study was to determine at what age children understand that gender is unchangeable and if that understanding is related to sex-typed preferences, as Kohlberg proposed. To accomplish this, the Sex Role Learning Index (SERLI) (Edelbrock & Sugawara, 1975) was administered. The SERLI includes one index of sex role knowledge, and two of sex-typed preferences. Children's occupational choices and individual toy choices were also included as indicators of sex-typed preferences. These measures and procedure were used to address the following research questions.

Research Questions

Question 1: At what age do children first demonstrate an understanding that gender is unchangeable?

Prediction 1: Consistent with prior research with appearance-reality distinction tasks, it was predicted that a developmental trend in the ability to pass the A-R Gender Test would emerge: It was expected that three-year-old children will not perform at above chance levels; four-year-old children would demonstrate some proficiency with the task; and five-year-old children would generally respond correctly.

Prediction 2: Because expressing transformation questions as appearance-reality distinction tasks is expected to avoid the methodological problems that inhibit children's performance on the Gender Constancy Interview, it was predicted that A-R Gender Test would be passed earlier than gender constancy transformation questions.

Prediction 3: It is assumed that gender as a category does not constitute a special type of identity. Thus, children should demonstrate fluency with identity transformations involving gender at the same time that they can solve transformation questions involving other objects. It is therefore predicted that performance on

the A-R Gender Test would be equivalent to performance on standard appearance-reality distinction tasks.

Question 2: Is the understanding that gender is unchangeable related to the development of sex-typed preferences?

Prediction 1: If understanding unchangeability leads children to positively value sex-appropriateness, children who pass the A-R Gender Test would be expected to have higher levels of sex-typed knowledge and preferences than children who do not pass the A-R Gender Test.

Prediction 2: Based upon the findings of previous studies, no relationship between performance on the Gender Constancy Interview and sex-typed preferences was expected.

Method

Subjects

Ninety-one children between the ages of three and six years were recruited from preschools and day care centres. Four subjects' data were subsequently eliminated: two four-year-olds had missing scores on more than two tasks; one four-year-old had extreme SERLI scores ($Z = -5.02$); and one three-year-old declined to complete the procedure. Thus, the final sample consisted of 87 children in three age groups: 28 three-year-olds (M = 42.89 mos. sd = 3.16 Range = 36-47 mos.; 14 males 14 females), 34 four-year-olds (M = 53.03 mos. sd = 3.33 Range = 48-59 mos.; 17 males and 17 females), and 25 five-year-olds (M = 67.73 mos. sd = 3.81 Range = 60-74 mos.; 13 males and 12 females). T-tests revealed no differences in average age of boys and girls within each group. Children had no apparent hearing or language problems, as indicated by the teachers.

Measures

Appearance-Reality Gender. The A-R Gender Test, developed for this study, was modeled on A-R distinction task procedures (Flavell, Flavell, & Green, 1983). The A-R Gender Test consists of two sets of four 13 X 18 cm. colour photographs. One set depicts a boy dressing up as a girl, and the other depicts a girl dressing up as a boy. The boy and girl are of approximately the same age and are pictured from the waist up, against the same background. The

transformation is demonstrated in four stages. The first photograph shows the child in sex-stereotyped clothing (the boy wears jeans and a t-shirt, the girl wears a dress). The second photograph shows the child in the first stages of dressing up (the boy is seen in the process of putting a dress on over his clothing, and the girl is seen pulling a t-shirt on over her dress). The third photograph depicts a change of hair (the boy is shown putting on a wig and the girl is shown tucking her hair up under a baseball cap). The last photograph shows the child fully dressed as an opposite sex child (the boy has on the dress and wig, and is holding a doll; the girl has on a t-shirt and a cap and is holding a baseball and bat). When shown the final photograph only (that of the cross-dressed child), all subjects in a group of adult raters (N = 10) identified the pictured child as the intended sex, confirming that the transformations are convincing ones.

The A-R Gender Test procedure was as follows: Children were first shown a picture of the crossed-dressed child and asked "What is this?". If subjects answered other than a gender label (e.g., said "a baseball player"), they were asked "Is it a boy or a girl?". The picture was then removed and the series of photographs depicting the transformation was laid out in front of the child, one picture at a time, with the following instructions: "See this boy? Well, look what he did one day. First he put on

this dress, then he put on this wig, and look, here he is all dressed up like a girl!" (subject's attention was drawn to the series by pointing to each picture in sequence). Children were given as much time as they want to examine the pictures. When the examiner was confident that each picture had been looked at, the series was then removed and the last picture only (that of the cross-dressed child) was re-presented with the test questions: (Appearance Question) "When you look at this with your eyes, right now, does it **look like** a girl, or does it **look like** a boy?" (Reality Question) "What is it **really and truly**? Is it **really and truly** a boy, or is it **really and truly** a girl?". The order in which alternative possible answers are presented were reversed for the picture of boy dressed as a girl and that of the girl dressed as a boy, to ensure that if children simply repeated the last choice presented, they answered incorrectly. Question order (Appearance first or Reality first) was counter-balanced across subjects. A photocopy of the pictures and administration procedure for the A-R Gender Test are included in Appendix A.

Gender Constancy. Gender constancy was assessed using the Gender Constancy Interview (Slaby & Frey, 1975). The interview consists of 10 questions concerning (in order) identity, stability, constancy (transformation), and motivation, as well as the counter-questions, as detailed earlier. A copy of the questions and procedure for

administering the Gender Constancy Interview is presented in Appendix B.

Appearance-Reality Distinction. The ability to distinguish appearance from reality was assessed with an Appearance-Reality Distinction Test (A-R Distinction Test). The A-R Distinction Test consisted of three items: 1) A sponge approximately 4 x 7 cm. that has been painted to look like a rock. The item was obtained from a professional supplier of theatre props and convincingly resembles a piece of granite; 2) A magnet, 3 x 3 cm, that was obtained from a novelty shop and looks very much like a chocolate candy; and 3) A piece of alabaster that was painted to look like an egg. Although the alabaster is exactly the same size and shape as an egg and has been painted with a eggshell coloured paint, it is considerably heavier. The objects were presented in the following manner, using the sponge painted as a rock as an example. First, the object was held in front of the children and they were asked "What is this?". After they responded the examiner said "Well, it sure looks like a rock, but let me show you something." The tester then squeezed the 'rock' and handed it to the child. "See? It looks like a rock, but it's really a sponge that someone painted. Do you know what a sponge is?" The tester and the child would then talk about sponges (e.g., how they are used in the bath, to wash dishes, how they are lighter than rocks, etc.,) until the child appeared to understand.

Then the object was taken back from the child, held up and the test questions were asked: (Appearance question) "Now, when you look at this with your eyes, right now, does it **look like** a rock or does it **look like** a sponge?" (Reality question) "What is it **really and truly**? Is it **really and truly** a sponge, or is it **really and truly** a rock?". Question order (Appearance first or Reality first) and the order in which objects were presented was counter-balanced across subjects.

The reliability of this procedure for assessing children's understanding of the appearance-reality distinction has not been conclusively established. However, Flavell et al. (1986) found children's performance to be resistant to training and stable when re-tested. Similarly, Astington and Gopnik (1988) found appearance-reality distinction task performance to be correlated with other measures of metacognitive abilities, such as understanding false beliefs and representational change. The administration procedure for the A-R Distinction Test is presented in Appendix C.

Sex Typing. Children's scores on the Sex Role Learning Index [SERLI] (Edelbrock & Sugawara, 1978) were used to determine level of sex-typing. The SERLI is an instrument that is frequently used to assess three related concepts: 1) Sex role discrimination (SRD) refers to children's knowledge of sex stereotypes and is indicated by the degree

to which a child's classification of sex-typed objects agrees with the sex stereotypes of those objects. SRD scores range from 0 (correctly assigned no items) to 20 (correctly assigned all items). The SRD section of the SERLI also yields a flexibility score, which reflects the number of times a child indicated that an item was appropriate for both males and females. Flexibility scores range from 0 (no activities considered for both) to 20 (all activities considered for both).

2) Sex role preference (SRP) refers to children's desire to adhere to societal sex role stereotypes and is the degree to which children's choices reflect traditional stereotypes. The SRP score is derived from the order in which children choose sex-typed activities from an array of items.

3) Sex role confirmation (SRC) is the degree to which children desire to adhere to their own perception of what is sex-appropriate. Instead of using traditional stereotypes to determine sex-appropriateness, the stereotyping of items is determined by the child's prior assessment of whether they are male or female activities. For example, when scoring SRP, wanting to be a soldier would be a sex-inappropriate choice for a girl. However, when scoring SRC, it would be a sex-inappropriate choice only if the girl had previously indicated that she believed that being a soldier is for men, but a <u>sex-appropriate</u> choice if she had indicated that it is for women. Thus, SRP is an index of

children's desire to adhere to traditional sex-typed activities, whereas SRC is an index of children's preference for what they believe to be sex-appropriate activities and avoids some of the difficulties arising from changing cultural stereotypes or children's different experiences.

The SRC score is derived from the order in which a child chooses sex-typed items from an array. Both the SRP and SRC sections yield a scale score of 20-80. A score of 20 indicates preference for sex-inappropriate activities and a score of 80 indicates complete preference for sex-appropriate activities. A score of 40 suggests no particular preference for male or female activities.

The SERLI items consist of black and white line drawings on 12 x 18 cm cards, balanced for size, detail, scale, and complexity. Items are divided into three sections: adult figures, child figures, and objects. In the figures sections, various sex-typed activities are depicted being performed by an adult or a child. The objects section contains twenty objects that are typically used by men or women, or boys or girls. The objects are used to assess SRD scores and the figures section is used to assess children's preferences for activities. Pictures are presented in a preset randomized order. For the SRD section, subjects are asked to determine whether or not each object is for males or females. For the SRP and SRC sections, children are presented with an array of items from the same-sex figures

section and asked to point one at a time to the activity they would most prefer to engage in. The child's choice is then removed and the question is then repeated until all pictures have been chosen, effectively producing a rank-order of the child's preferences (see Appendix D for pictures).

As mentioned above, the SERLI yields three indices of sex-typing. The SRD section is scored as the percentage of agreement between children's classification of objects and the objects' sex-type, the SRP score is determined by the child's choice of stereotypical items, and the SRC score is determined from the child's own classification of whether the item is masculine or feminine. For each of the SRC and SRP scores, scoring is based upon the order in which items are chosen and the probability of making that choice by chance. The scoring system takes into account the number of possible choices (both sex-typed and non-sex-typed) that a child has when assigning weight to the choice he or she makes. For example, if a boy's first choice out of 10 items (of which five are 'male' five are 'female') is sex-inappropriate, it is of greater significance than if the child's choice is sex-inappropriate when no other sex-appropriate choices remain. Thus, the scores obtained with the SERLI probability system of scoring reflect the relative strength of the child's preferences given the alternative choices, taking into account the probability of selecting an

item by chance alone. It is therefore a relatively sensitive index of sex-typing.

Toy Preferences. Toy preferences were assessed as an additional index of sex-typed preferences. Children were asked to indicate which toys they prefer to play with, from an array of ten 5 X 7" line drawings of toys (five 'male', five 'female'). The items were developed for this study based upon previous research of children's toy preferences, both in free play and forced-choice situations (Delucia, 1963; Fagot, 1985). Care was taken to equate pictures for size of item, level of detail, and type of activity. Thus, both a 'female' and 'male' toy was depicted for each of the following activities: fine motor play (crayons and paints / small building blocks), gross motor play (skipping rope / wagon), dress-up (dress, purse, and high-heeled shoes / pirate costume), doll-type play (baby doll / robot doll), and playing at adult roles (tea set / tool set). Twenty preschool children were asked to sort the items on the basis of whether they were 'boy's or 'girl's toys,' to confirm the intended sex stereotype of the toys. In general, agreement on whether items were for boys or girls was high for nine of the ten items (M Proportion Agreement across items = .86, sd.=8.9). The tenth item, the robot doll, was labelled as being for boys by only 13 of the 20 children (65%). Given the difficulty of trying to replace the robot with a different doll-like toy for boys, it was decided to retain

this item.

Toy items were included as an extra preference measure to compensate for a possible problem with the SERLI preference items. Although research with the SERLI has indicated that the preference scale is reliable (test-retest r = .90 and .84 for child and adult figures, respectively) and moderately correlated with other indices of sex-typed preferences (r = .45 and .38 for child and adult figures, respectively; Edelbrock & Sugawara, 1978), the objects depicted for each sex may influence sex-typed choices. The choices for the boys include four occupations (fire fighter, police officer, doctor, soldier), three adult activities or chores (digging with a shovel, sawing, hammering) and three games (boxing, baseball, car play). The girl's choices include two adult occupations, one picture depicting a teacher, and seven pictures depicting the role of home maker (feeding baby, pouring juice, baking, sweeping, cooking, washing dishes, ironing). There are also two choices which could be considered adult activities or chores (brushing hair, sewing). The girl's choices include no games, fewer occupational choices, and domestic chores are most frequently represented.

Although play at adult roles is definitely a source of enjoyment for children of both sexes, the different choices available in the SERLI preference section may affect the degree to which boys and girls demonstrate a preference for

sex-typed activities. The 10 new toy choices were included to offer equal activity choices in the measurement of children's preferences. Because the SERLI scoring method is based upon the order in which children choose various items and requires only that the available items be sex-typed, it is possible to administer and score the new toy items in the same manner. Once it was established whether items were stereotypically for males or females, the items were randomly assigned to an array for presentation, using the same assignment procedure as was used for the SERLI. Similarly, the ten items were presented with the SRD portion of the SERLI items, to obtain children's own perception of the sex-typing of the items. In this manner, the toy items can be scored for SRD, SRP, and SRC, as the SERLI items are, without compromising the SERLI test administration or scoring. A photocopy of the toy items is presented in Appendix E.

Occupational Choices. Children were asked what they would like to be when they grow up. This question was posed in casual conversation with children. It was included because there is some evidence that children are sensitive to changes in typical adult occupations and, even when they have sex-typed preferences for toys and activities, they can hold very non-sex-typed occupational preferences (Weisner & Wilson-Mitchell, 1990). Occupational choices were included as an additional indicator of sex-typed preferences.

Children's responses were coded as 1) traditionally male, 2) traditionally female, 3) other (e.g., non-human), or 4) no answer.

Procedure

Testing took place in the children's preschool or day care. Children were individually tested by a female examiner at a small table in a separate area of the school. Each test was administered according to the procedure outlined in Appendices C through E, and children's responses were recorded verbatim. To establish the consistency of test administration, 36% of the children (31) were tested by a second trained examiner who was blind to the hypotheses of the study. The second examiner tested at least five girls and five boys at each age level.

Test presentation order was counterbalanced to accomplish the following constraints. First, to examine whether or not the first question influences subsequent responding on A-R Tasks, the appearance question preceded the reality question half the time. Second, because the A-R Gender Test and Gender Constancy Interview deal with similar transformations, there may be carry-over effects in responses. Therefore, the A-R Gender Test and Gender Constancy Interview did not occur consecutively, and the A-R Gender Test preceded that Gender Constancy Interview half the time.

Children were tested in one session, 15 to 20 minutes in length, after which they were thanked for their help and given stickers and a personalized "Thank You" card for their participation.

Children's protocols were scored by the author. Thirteen protocols (15%) were randomly selected and scored by a second person who had not tested any subjects. For the A-R Distinction Test, A-R Gender Test, and the Gender Constancy Interview, for which most items were scored either pass or fail, Kappa Coefficients (Cohen, 1960) were calculated for each task. Obtained coefficients were .95, .98, and .99, respectively. Scoring the SERLI required the transcription of numbers and numerical calculations. The scores were checked by the second scorer and found to be 99% accurate (4 errors out of 949 scores).

Results

Preliminary Analyses

Before addressing the main hypotheses, tests of experimenter and order effects were conducted, followed by analyses of each measure. Analyses of the main research questions were then conducted.

Tests of Experimenter and Order Effects

Thirty-one subjects (36% of final sample) were tested by a second experimenter who was blind to the hypotheses of the study. The second experimenter tested at least five boys and five girls at each age level. A t-test confirmed that there was no difference in the average age (in months) of children tested by each experimenter, M = 54.2, sd.= 10.4, M = 53.7, sd. = 10.4, t (85) = .22, p< .82. Chi-Square tests indicated that, for the Appearance-Reality Distinction Task and the Appearance-Reality Gender Task, the proportion of children passing and failing each measure did not differ by experimenter, χ^2 (1, N = 87) = .66, p<.42; χ^2 (1, N = 87) = 1.9, p<.82, respectively. Similarly, there was no difference in the average gender constancy stage achieved by children tested by each experimenter. T-tests of the difference between the average score obtained on the eight SERLI subscales revealed one significant difference. Children tested by the author had higher sex role confirmation scores for adult roles (SRCAD) (M = 61.28, sd =

11.46, N = 56) than did children tested by the second experimenter (M = 53.52, sd = 9.35, N = 31), t (83) = 3.13, p<. 003. Subsequent examination of the SRCAD scores of children tested by the author indicated that they did not differ systematically by age or sex. The means and standard deviations for the eight SERLI subscales by experimenter are presented in Table 1.

Presentation order was balanced to ensure that half the subjects received appearance questions before reality questions and half the subjects received the Gender Constancy Interview before Appearance-Reality Gender Task. Chi-Square tests indicated that children who received appearance questions first or reality questions first did not differ in their rates of passing or failing the Appearance-Reality Distinction Task, χ^2 (1, N = 87) = .34, p< .56, or Appearance-Reality Gender Task, χ^2 (1, N = 87) = .35, p< .55. Similarly, children who received the Gender Constancy Interview first did not differ from those who received the Appearance-Reality Gender Task first in the rates of passing or failing Gender Constancy, χ^2 (1, N = 87) = 1.45, p<. 23, or Appearance-Reality Gender Task, χ^2 (1, N = 87) = 2.2, p< .14.

Another possible outcome of presentation order is that of differences in the type of errors children make when answering appearance-reality questions. As described previously, when answering appearance-reality type questions

Table 1

Means and Standard Deviations of SERLI Subscale Scores by Experimenter

OWN-SEX ROLE KNOWLEDGE:

	N	M	SD	t	DF	p
EXP 1	56	94.46	(7.8)	1.30	(85)	.19
EXP 2	31	91.61	(12.7)			

OPP-SEX ROLE KNOWLEDGE:

	N	M	SD	t	DF	p
EXP 1	56	82.50	(16.7)	.49	(85)	.62
EXP 2	31	80.65	(16.9)			

SEX ROLE PREFERENCE (CHILD ACTIVITIES):

	N	M	SD	t	DF	p
EXP 1	56	59.82	(11.9)	.54	(85)	.59
EXP 2	31	58.32	(13.0)			

SEX ROLE PREFERENCE (ADULT ACTIVITIES):

	N	M	SD	t	DF	p
EXP 1	56	61.50	(10.7)	1.34	(85)	.19
EXP 2	31	58.00	(13.4)			

SEX ROLE CONFIRMATION (CHILD ACTIVITIES):

	N	M	SD	t	DF	p
EXP 1	56	57.88	(11.9)	1.04	(83)	.30
EXP 2	31	54.97	(12.8)			

(table 1 continues)

Table 1 (Continued)

SEX ROLE CONFIRMATION (ADULT ACTIVITIES):

	N	M	SD	t	DF	p
EXP 1	56	61.27	(11.5)	3.13	(83)	.002
EXP 2	31	53.55	(9.4)			

SEX ROLE PREFERENCE (TOYS):

	N	M	SD	t	DF	p
EXP 1	56	62.84	(11.6)	1.01	(85)	.31
EXP 2	31	60.00	(14.1)			

SEX ROLE CONFIRMATION (TOYS):

	N	M	SD	t	DF	p
EXP 1	56	62.28	(12.4)	1.31	(82)	.19
EXP 2	31	58.29	(15.2)			

children could err in two ways. First, they could make a reality-based response to the appearance question and say that the object looks like what it really is (e.g., that the rock/sponge looks like a sponge). This type of error is called a realism error. Second, children could make an appearance-based response to the reality question, saying that the object really is what it looks like (e.g., that the rock/sponge really is a rock). This is called a phenomenism error. Children could make both errors on each object, and therefore the error scores were neither mutually exclusive nor reciprocal.

To determine if question order affected the type of errors children made, t-tests of the difference between the number of phenomenism and realism errors made by children who received the appearance question or reality question first were conducted. Children who received the appearance questions first did not make more phenomenism errors ($\underline{M}$ = .864, sd. = 1.0) than did children who received the reality questions first ($\underline{M}$ = .721, sd. = 1.2) $\underline{t}$ (85) = .59, $\underline{p}$< .56. Similarly, children who received the appearance questions first did not make more realism errors ($\underline{M}$ = 1.46, sd. = 1.2) than did children who received the reality questions first ($\underline{M}$ = 1.44, sd. = 1.3) $\underline{t}$ (85) = .04, $\underline{p}$< .96. Thus, no difference was found in the average number of each type of error made on the Appearance-Reality Distinction Task or the Appearance-Reality Gender task as a function of question order.

To determine if receiving the Gender Constancy Interview first or the Appearance-Reality Gender Task first affected the type of errors children made, a t-test of the difference between phenomenism and realism scores was conducted. It was found that children who received the Gender Constancy Interview first made more realism than phenomenism errors when responding to the Appearance-Reality Gender Task ($\underline{t}$ (85) = 2.18, $\underline{p}$< .05). Although this difference was not significant with the Bonferroni correction for multiple t-tests, it was further investigated

to determine what systematic effects it may have had upon children's responding.

Uncorrected t-tests indicated that this difference was significant for four-year-old girls only. The eight four-year-old girls who received the Gender Constancy Interview first made more realism errors than phenomenism errors. However, further investigation revealed that the eight four-year-old girls who received the Gender Constancy Interview first did not differ from four-year-old boys in the number of realism errors made, nor did they differ from the rest of the sample combined. Thus, the significant difference between realism and phenomenism errors within the four-year-old girls who received different presentation orders may be a result of the very low number of phenomenism errors four-year-old girls made overall, and on this task in particular ($\underline{M}$ = .22 vs. $\underline{M}$ = .50 for the entire sample). We were satisfied that there were no important differences in children's performance due to experimenter or order effects. Therefore, subjects were collapsed across presentation order and experimenter for subsequent analyses.

Tests of the Measures

Preliminary tests of the measures were conducted to screen for differences in children's performance as a function of sex or specific objects tested and to ensure that performance on the measures was similar to that found

in other studies.

1) Appearance-Reality Distinction Test. Subjects' performance on this task was examined in three areas: all objects combined, individual objects, and error patterns. Each is presented separately.

All Objects Combined: The Appearance-Reality Distinction Test (A-R Distinction Test) included questions about three objects (rock/sponge, chocolate/magnet, egg/stone). To have been credited with passing an object, children had to correctly answer two questions (what the object looked like and what it really was), thus, the probability of passing any single object by chance alone was .25. The probability of passing all three objects by chance was .016. The criterion for having been credited with passing the A-R Distinction Test was two out of three objects correct. Chi-square analysis indicated that there was an effect of age group on passing rates, with 39% of the three-year-olds, 62% of the four-year-olds, and 88% of the five-year-olds passing the A-R Distinction Test, χ^2 (2, N = 87) = 13.32, $p< .01$. The number of children passing the A-R Distinction Test at each age level using the criterion of passing two out of three objects is presented in Table 2. A Chi-Square test indicated that there was no difference between the number of boys and girls passing the A-R Distinction Test, χ^2 (1, N = 87) = 1.41, $p< .24$.

Table 2

Number and Percentage of Children Passing Appearance-Reality Distinction Test by Age Group

	AGE GROUP			
	3 yrs	4 yrs	5 yrs	TOTAL
PASS	11 (39%)	21 (62%)	22 (88%)	54 (62%)
FAIL	17 (61%)	13 (38%)	3 (12%)	33 (38%)
TOTAL	28	34	25	N = 87

Individual Objects: Chi-Square tests of the number of children passing each A-R Distinction Test object were conducted separately for boys and girls. Within the three- and four-year-old age groups, there were no differences in the number of boys and girls passing any one of the A-R Distinction Test objects. Within the five-year-old group, however, the low number of children failing objects precluded Chi-Square analysis. Only three five-year-old children failed to pass two out of three objects, and those children's errors were distributed across all three objects (all three failed the chocolate/magnet and the egg/stone, two also failed the rock/sponge). Therefore, children in different age groups were combined to examine children's responses on individual objects. Chi-Square tests indicated

no sex differences in the number of children passing the rock/sponge, χ^2 (1, $\underline{N}$ = 87) = .98, $\underline{p}$< .32; chocolate /magnet, χ^2 (1, $\underline{N}$ = 87) = .92, $\underline{p}$< .34; or the egg/stone, χ^2 (1, $\underline{N}$ = 87) = .01, $\underline{p}$< .92.

Error Analyses: As described earlier, children can make realism or phenomenism errors when responding to appearance-reality distinction questions. To determine if children made different types of errors at different ages, an Age group (3) by Error type (2) ANOVA was conducted. The main effect of Age group was significant, $\underline{F}$ (2, 84) = 12.02, $\underline{p}$< .001. Examination of the main effect of age (Newman-Keuls Test) indicated that the average number of errors made was significantly different between each age group (of three possible errors each, $\underline{M}$ = 1.89, sd. = 1.2; $\underline{M}$ = 1.26, sd. = 1.16; $\underline{M}$ = .40, sd. = .91; for three-, four-, and five-year-olds, respectively). The main effect of Error type was also significant, $\underline{F}$ (1, 84) = 8.48, $\underline{p}$< .005, with children making more realism errors ($\underline{M}$ = .83, sd. = 1.06) than phenomenism errors ($\underline{M}$ = .39, sd. = .72). The Age group by Error type interaction was not significant. Further analyses of the children's errors will be discussed in detail with the main analyses.

Summary of A-R Distinction Test. Children's performance on the A-R Distinction Test revealed a clear developmental trend in the number of children meeting the minimum passing criterion of two out of three objects.

There were no sex differences in the number of children passing overall or passing each object separately. Children consistently made more realism than phenomenism errors, and the number of errors decreased with age. Responses to the A-R Distinction Test in this study replicate the findings of other appearance-reality distinction studies. Specifically, the proportion of children passing in each age group and the higher incidence of realism errors are consistent with the rates found with similar appearance-reality distinction tasks (e.g., Flavell, Flavell, & Green 1983; Flavell, Green, & Flavell, 1986).

2) Appearance-Reality Gender Task. Subject's performance on the A-R Gender Test was examined in three areas: all objects combined, individual objects, and error patterns. Each is presented separately.

All Objects Combined: The Appearance-Reality Gender Task (A-R Gender Test) consisted of two tasks (picture series of a boy dressed as a girl and of a girl dressed as a boy). In order to have been credited with passing one of the tasks, children had to have correctly answered two questions regarding what the pictured child looked like (appearance questions) and what it really was (reality question).

For any single subject, the probability of passing any one picture by chance alone is. 25, and of responding

correctly on both pictures is .06. Although children passed at above-chance levels (i.e., more than 6% of children in each age group), the difficulty level of this task resulted in very low rates of occurrence: only 18% of three-year-olds, 35% of the four-year-olds, and 56% five-year-olds passed both pictures of the A-R Gender Test. A Chi-square test indicated that these proportions differed by age, χ^2 (2, $\underline{N}$ = 87) = 8.38, $\underline{p}$< .01.

Sex differences in the number of children passing and failing the A-R Gender Test could not be examined within each age group because few three-year-old children passed the A-R Gender Test, which resulted in cells with small observed frequencies thus violating an assumption of the Chi-Square test. However, when age groups were combined, a Chi-Square test revealed no significant difference between the number of boys and girls passing or failing the A-R Gender Test overall, χ^2 (1, $\underline{N}$ = 87) = 2.71, $\underline{p}$< .10.

Individual Pictures: Children's performance on the pictures of the boy dressed as girl and of the girl dressed as boy was examined separately to determine if there was a difference in the number of boys or girls passing or failing each picture. Again, the small number of children passing the A-R Gender Test at three and four years of age necessitated collapsing across age groups to examine sex effects and collapsing across sex to examine age effects.

Chi-Square tests indicated no difference in the number

of boys and girls passing the picture of the girl as a boy. There was, however, a difference for the picture of the boy dressed as a girl, χ^2 (1, $\underline{N}$ = 87) = 4.21 $\underline{p}$ <.04. For this picture, 28 (65%) of the girls passed whereas only 19 (43%) of the boys passed. A closer examination of the differences in boy's and girl's responding on this task is presented with error analyses, below. The number of children meeting the criterion for passing the A-R Gender Test, by age group and sex, is presented in Table 3.

Error Analyses: As described earlier, children can make realism or phenomenism errors when responding to the A-R Gender Test. Children made a similar number of realism and phenomenism errors ($\underline{M}$ =.621, sd. = .73 vs. $\underline{M}$ =.420, sd. = .71) $\underline{t}$ (86) = 1.86, $\underline{p}$< .07. Age and type of errors could not be combined as variables in an ANOVA because none of the five-year-old children made phenomenism errors. Therefore, phenomenism and realism errors were examined separately. Phenomenism errors declined with age. Fifty-seven percent of the three-year-olds made at least one phenomenism error, compared with twenty-four percent of the four-year-olds, $\underline{Z}$ = 2.66, $\underline{p}$< .01. The five-year-olds made no phenomenism errors. Interestingly, realism errors did not decline significantly with age. A Chi-Square test of the number of children making at least one realism error indicated no effect of age (13%, 17%, and 11% for three-, four-, and five-year-olds, respectively), χ^2 (2, $\underline{N}$ = 87) = .216,

Table 3

Number of Boys and Girls Passing Appearance-Reality Gender by Age Group and Sex

	AGE GROUP						
	3 yrs		4 yrs		5 yrs		TOTAL
	BOYS	GIRLS	BOYS	GIRLS	BOYS	GIRLS	TOTAL
PASS	2	3	4	8	6	8	31
%	(14%)	(21%)	(24%)	(47%)	(46%)	(67%)	(36%)
FAIL	12	11	13	9	7	4	56
%	(86%)	(79%)	(76%)	(53%)	(54%)	(33%)	(64%)
N =	28		34		25		87

$p < .90$. Thus, on the A-R Gender Test the percentage of children making phenomenism errors decreased sharply with age whereas realism errors did not decline as markedly. These differences, as well as further analyses of sex by object effects, will be discussed in detail with the main analyses.

Summary of A-R Gender Test. Children's performance on the A-R Gender Test revealed a clear developmental trend in the number of children who passed both pictures in the A-R Gender Test. At each age the number of children passing was above the level expected by chance alone but was too small to suggest proficiency in the three- and four-year-old groups. Although girls and boys did not differ in their performance on the picture of a girl dressed as a boy, significantly more girls than boys passed the picture of a boy dressed as a girl. Overall, children did not make significantly more phenomenism errors than realism errors on A-R Gender Test. In general, phenomenism errors declined with age whereas realism errors did not decline as markedly.

3) The Gender Constancy Interview. The Gender Constancy Interview contains three levels of questions that have been demonstrated to form hierarchical stages. If subjects correctly answered all the identity questions, they were credited with having achieved Stage 1. If they achieved Stage 1 and correctly answered all the stability questions, they were credited with Stage 2. If they achieved Stage 2 and answered the constancy questions correctly, they were credited with having achieved Stage 3. Children could also fail to achieve any stage, in which case they were credited with Stage 0. In the present sample, 82 children (94%) performed in one of the three expected scale patterns and

Table 4

Number of Children Achieving Gender Constancy Stage by Age Group

AGE	STAGE 0	STAGE 1	STAGE 2	STAGE 3	TOTAL
3 Yrs	4 (14%)	3 (11%)	8 (29%)	13 (46%)	28
4 Yrs	1 (3%)	4 (12%)	18 (53%)	11 (32%)	34
5 Yrs	0	1 (4%)	10 (40%)	14 (56%)	25
TOTAL	5 (6%)	8 (9%)	36 (41%)	38 (44%)	N=87

only five children (6%) did not. The number of children achieving Gender Constancy Stages in each age group is presented in Table 4. Children in the sample were of a relatively advanced stage of Gender Constancy (sample M = 2.23, sd= .84). In each age group more than 75% of the children were in Stage 2 or Stage 3. A 2 (Sex) x 3 (Age group) ANOVA of the average Gender Constancy Interview scores achieved revealed no significant main effects or interaction.

Summary of Gender Constancy Measure. Performance on the Gender constancy Interview indicated that few four- and five-year-olds failed to achieve Stage 2. The average age

at which children achieved each stage and the relatively advanced Gender Constancy Interview scores was similar to that found by others (e.g., Edelbrock & Sugawara, 1978; Taylor & Carter, 1987; Martin & Little, 1990).

4) SERLI Sex Typing Index.

Knowledge Scores: Preliminary analyses revealed that sex role knowledge scores were negatively skewed, indicating that children had a substantial amount of sex role knowledge. However, when the following analyses were conducted on raw scores and on scores subject to loglinear transformations, the results did not differ. Furthermore, Tabachnik and Fidel (1989, p. 71) have suggested that ANOVA is robust to violations of the normality assumption when degrees of freedom exceed 200. Therefore, for ease of interpretation the following analyses are of the untransformed data.

To examine possible age and sex effects on knowledge scores, a 2 (Sex) x 2 (Score-type: own or opposite) by 3 (Age group) ANOVA of children's knowledge scores was conducted. Results indicated a significant main effect of Score type, F (1,81) = 40.73, $p < .001$, a main effect of Age group, F (2,81) = 12.66, $p < .001$, and an Age by Score type interaction, F (2,81) = 3.34 $p < .04$. Post-hoc comparisons (Newman-Keuls Test) indicated that children in the three-year-old group had significantly lower own-sex knowledge

(Max. = 100: M = 89.64, sd. = 12.9) scores than did children in the five-year-old group (M = 96.8, sd. = 6.27). Three-year-olds also had significantly lower opposite-sex knowledge scores (M = 71.43, sd. = 20.68) than did either the four-year-olds (M = 85.0, sd. = 11.9) or the five-year-olds (M = 89.2, sd. = 11.2). Examination of the means of own- and other-sex knowledge scores indicated that, overall, knowledge of own sex (M = 93.45, sd. = 9.86) was greater than knowledge of other sex (M = 81.84, sd. = 16.7).

Preference Scores: To examine possible age and sex effects in preference scores, a 2 (Sex) x 2 (Preference score type: SRP SRC) x 3 (Age group) x 3 (Activity type: Child, Adult, or Toy) ANOVA of children's preference scores was conducted. All main effects were significant as was the interaction of age by preference score. The means and standard deviations for these variables are presented in Table 5. The source table for the ANOVA is presented in Appendix F.

Examination of the main effect of age (Newman-Keuls Test) indicated that three-year-olds had lower overall SERLI scores than did four- or five-year-olds (M = 56.34, sd. = 7.61; 61.19, sd. = 8.32; 61.44, sd. = 9.19, respectively).

Table 5

Means and Standard Deviations of Variables in ANOVA of SERLI Scores by Age, Sex, Task Type and Score Type

AVERAGE SERLI SCORES

BY AGE:			
	3 YRS	4 YRS	5 YRS
	56.34 (7.60)	61.19 (8.32)	61.44 (9.19)
BY SEX:			
	BOYS	GIRLS	
	61.48 (7.25)	57.88 (9.50)	
BY ACTIVITY TYPE:			
	CHILD	ADULT	TOYS
	58.18 (11.37)	59.51 (10.55)	61.40 (12.51)
BY SCORE TYPE:			
	SRP	SRC	
	60.45 (9.08)	58.54 (9.09)	
AGE X SCORE TYPE INTERACTION:			
	3 YRS	4 YRS	5 YRS
SRP	58.07 (7.84)	61.50 (9.12)	61.71 (10.1)
SRC	54.02 (8.34)	60.47 (8.87)	61.00 (8.62)

Examination of the means for the main effect of sex indicated that the boys had significantly higher overall scores than did girls ($\underline{M}$ = 61.47, sd. = 7.26 vs. 57.87, sd. = 9.50). Comparison of the means (Newman-Keuls Test) for the main effect of activity type (Child, Adult, and Toy preference scores, $\underline{M}$ = 58.18, sd. = 11.37; $\underline{M}$ = 59.51, sd. = 10.55; and $\underline{M}$ = 61.4, sd. = 12.31, respectively) indicated a significant difference between child and toy preferences only.

The interaction between age group and preference score was examined with planned comparisons of the average preference score at each age. Comparison of the means (Newman-Keuls Test) indicated that three-year-old children had significantly lower SRC scores ($\underline{M}$ = 54.02, sd. = 8.34) than did the four-year-old ($\underline{M}$ = 60.47, sd. = 8.87) and five-year-old ($\underline{M}$ = 61.0, sd. = 8.62) children. SRP scores did not differ. Interestingly, three-year-old children's SRC scores ($\underline{M}$ = 54.02, sd. = 8.3) were lower than their SRP scores ($\underline{M}$ = 58.07, sd. = 7.8), $\underline{t}$ (27) = -5.16, $\underline{p}$< .001. This is a curious discrepancy given the lower overall knowledge scores of the three-year-olds. It suggests that the three-year-old children preferred stereotypical sex-appropriate activities (high SRP scores) without necessarily knowing that they were sex-appropriate (low knowledge scores). Furthermore, although they adhered to traditional sex roles, they did not appear to feel compelled to be sex-

appropriate: they did not indicate a preference for those activities that they themselves had determined were sex-appropriate (low SRC scores).

Because the girls' adult activity choices on the SERLI are restricted in terms of variety, ten toy items balanced for type of play or activity involved had been included as an extra index of sex-typed preferences. T-tests of boys' and girls' average toy preference scores revealed no significant difference in their preference for sex-typed toys overall ($\underline{M}$ = 63.1, sd = 11.0 vs. $\underline{M}$ = 59.2, sd = 13.3 for boys and girls, respectively) or when examined separately by age group. Thus, on the SERLI, girls appear to have lower sex-typed preferences. However, when toy choices were offered that provide a variety of activity choices greater male preference for sex-typed activities was not found.

Summary of SERLI Scores: Preliminary analyses of the SERLI scores indicated that three-year-old children had significantly less knowledge of their own sex than did older children and that all children had more knowledge of their own sex than of the opposite sex. Boys and girls did not differ in their knowledge of sex roles. Children's preference scores differed with age as well: three-year-old children had higher SRP than SRC scores. Four- and five-year-old children did not differ in their SRP or SRC indices. As has been found elsewhere, boys had higher SERLI

preference scores than girls. However, boys and girls did not differ in their preference for sex-typed toys, suggesting that girls' lower preference scores may reflect the less attractive choices offered rather than overall lower sex role preferences. For both boys and girls, preference scores for sex-typed toys were higher than preference scores for either sex-typed child or adult activities.

The performance of children in the present sample resembles that of children in previous studies with the SERLI. Specifically, the higher own-sex vs. other-sex knowledge, the higher boys' preference scores, and the general increase in sex-typed knowledge with age, are similar to the findings of Edelbrock and Sugawara (1978).

5) Occupational Choices. Children's choice of adult occupation was scored as sex-typed, non-sex-typed, other, or no response. Not surprisingly, three-year-olds had difficulty with the question of what they wanted to be when grown up and frequently responded with non-human choices (i.e., a duck, a tractor) or could not answer. Four- and five-year-olds made more real-life choices. Chi-Square tests indicated an effect of age, χ^2 (2, N = 87) = 6.98, p< .03, with 42%, 54%, and 72% of the three-, four-, and five-year-old children making sex-typed choices. There were sex differences in children's preference for sex-typed

occupations. Although girls preferred male and female sex-typed occupations equally, boys consistently preferred male occupations, χ^2 (1) = 19.11; $p < .001$. Children who made sex-typed occupational choices had higher SERLI Sex Role Preference scores (M = 63.02, sd. = 8.6 vs. M = 57.57, sd. = 8.8), t (85) = 2.91, $p < .005$, and Sex Role Confirmation scores (M = 60.65, sd. = 9.1 vs. M = 56.18, sd. = 8.7), t (85) = 2.35, $p < .03$, although they did not differ in their Sex Role Knowledge scores (M = 89.24, sd. = 9.3 vs. M = 85.84, sd. = 11.8), t (85) = 1.5, $p < .14$.

Summary of Preliminary Analyses

Preliminary analyses of children's performance on the A-R Distinction Test, Gender Constancy Interview, and the SERLI indicated that performance on the measures was consistent with that found in previous studies. Next, the main research questions were addressed.

Tests of the Research Questions

Q. 1) At what age do children understand that gender is unchangeable?

Predictions 1, 2, and 3

There were three predictions associated with the first question. The first prediction was that a developmental trend in the ability to pass the A-R Gender Test would be

evident. Three-year-old children were not expected to pass the A-R Gender Test, four-year-olds were expected to show some competence with the A-R Gender Test, and five-year-olds were expected to pass the A-R Gender Test. As indicated in the preliminary analyses, an age-related trend in the ability to pass the A-R Gender Test was evident. However, although children passed at above-chance levels within each age group, only 18%, 35%, and 56% of the three-, four-, and five-year-olds passed. The second prediction was that the ability to pass the A-R Gender Test would be unrelated to and precede achieving Stage 3 of the Gender Constancy Interview. Pearson correlations indicated that A-R Gender Test scores were unrelated to Gender Constancy Stage, r (N = 87) = .11, p<.46. However, the prediction that the ability to pass the A-R Gender Test would precede Gender Constancy Stage 3 was not supported. The number of children in each age group who achieved Stage 3 of the Gender Constancy Measure but did or did not pass the A-R Gender Test, is presented in Table 6. Contrary to prediction, of those children who achieved Stage 3, fewer three-year-olds passed the A-R Gender Test (23%) than failed (77%), z = -2.75, p< .001, slightly but not significantly fewer four-year-olds passed the A-R Gender Test (36%) than failed (64%), z = -1.31, p< .09, and an equal percentage of five-year-olds passed (50%) and failed (50%) the A-R Gender Test.

Table 6

Number of Children Achieving Stage 3 of Gender Constancy Interview who Passed and Failed A-R Gender Test, by Age Group

	AGE GROUP		
	3 YRS	4 YRS	5 YRS
PASSED A-R GENDER	3 (23%)	4 (36%)	7 (50%)
DID NOT PASS A-R GENDER	10 (77%)	7 (64%)	7 (50%)
TOTAL ACHIEVING STG 3 WITHIN AGE GROUP	13	11	14
PERCENT OF TOTAL AGE GROUP	46%	32%	56%

The final prediction was that performance on the A-R Gender Test and the A-R Distinction Test would be correlated. There was a small positive correlation between A-R Gender Test scores and A-R Distinction Test scores, r (N = 87) =.24, p< .01. However, as shown in Table 7, there were substantial base rate differences in the number of children passing at each age level. More children passed the A-R Distinction Test than the A-R Gender Test at three years (39% vs. 18%, z = 1.65, p<. 05), four years (62% vs. 35%, z = 2.27, p< .01) and at five years of age (88% vs. 56%, z = 2.44, p< .05). This suggests that gender as a category may constitute a special case of identity. This possibility is further discussed with the error analyses, below.

Although these findings partially address the three predictions, the general question of when children understand that gender is unchangeable is best approached by examining the developmental sequence by which competency emerges on the A-R Gender Test, A-R Distinction Test, and the Gender Constancy Interview.

Developmental Sequence: Weinraub et al. (1984) have presented a procedure for establishing developmental trends in task performance based upon two features: age of onset and age of majority. Age of onset is the age at which the ability to pass a given task is reliably demonstrated. To determine the age of onset, the number of children meeting

Table 7

Proportion of Children Passing Age of Majority for A-R Distinction Test, A-R Gender Test, and The Gender Constancy Interview

	AGE GROUP		
TASK	3 YRS	4 YRS	5 YRS
A-RD	39 % Passed p<.005	62 % Passed p<.0001 **MAJORITY** *	88 % Passed p<.0001 **MAJORITY** *
A-RG	18 % Passed p<.016	35 % Passed p<.0001	56 % Passed p<.0001 **MAJORITY** *
G. C.	46 % Passed	32 % Passed	56 % Passed **MAJORITY** *

Note: p value indicates the probability that the percentage of children passed by chance alone; **MAJORITY** indicates that the percentage of children passing meets the criterion for Age of Majority

the criteria for passing the task is tested against the number of children expected to have passed by chance. The binomial approximation of the normal curve was used to test the hypothesis p = X, where X is the proportion of children passing the task, and p is the probability of having passed the criteria by chance. For the A-R Gender Test, the probability of correctly guessing both questions for both pictures is .063. For A-R Distinction Test, the probability of guessing two out of three objects is .152. Stage 3 of the Gender Constancy Interview is achieved by correctly answering 18 pass/fail questions, resulting in an extremely low probability of achieving Stage 3 by chance alone (p=.0000038). Because of the low probability of passing each of these tasks by chance alone, even a small number of children passing would be sufficient to meet Weinraub's criteria for age of onset, which would not be a meaningful indicator of that age level's proficiency with the task. The second criterion, age of majority, yields a more appropriate index of proficiency.

Age of majority is the age at which most children within an age group reliably pass a given task and is described as follows. Within an age group, the number of children expected to respond correctly by chance alone is p x N. Because chance-level responding is presumed to be randomly distributed, 50% of the sample could also be expected to contain half the number of chance respondents.

Therefore, the criterion for majority is 50% + (p x N)/2, and the age of majority is the youngest age at which this number of children pass the task. For A-R Gender Test, majority is 50% + (.0625 x N)/2, at each age group. For A-R Distinction Test, the majority is 50% + (.1517 x N)/2, at each age group. For the Gender Constancy stage 3, majority is 50% + (.0000038 X N)/2, which essentially amounts to 50% at each age group.

As can be seen in Table 7, age of majority emerged in a clear developmental trend: the A-R Distinction Test was not passed by the majority of children until four years of age. The majority of children did not pass A-R Gender Test until five years of age. Similarly, Gender Constancy Stage 3 was not passed by the majority of children until five years of age, although a substantial number passed at three years of age.

This developmental trend suggests that these tasks may be hierarchical in nature. Children's individual patterns of responses were examined to determine whether or not they conformed to a Guttman scale. There are four possible patterns of passing the three tasks which would be expected if the tasks formed a hierarchical scale: a child could fail all three tasks; could pass the A-R Distinction Test but fail the A-R Gender Test and the Gender Constancy Interview Stage 3; could pass the A-R Distinction Test and the A-R Gender Test but fail Gender Constancy Interview Stage 3; or

could pass all three. Any other combination would constitute an error pattern. If the tasks form a perfect scale, all children would conform to one of the expected patterns. In the present sample, only 64% of the children conformed to an acceptable pattern and 46% made error patterns, yielding a Guttman Reproducibility Coefficient of .76. Coefficients of at least .85 are recommended before tasks can be considered to have scale-like properties (Nie & Hull 1975). Thus, although performance on the present tasks emerged in a clear developmental pattern in terms of age at which the majority of children passed, the individual patterns indicated that the tasks are not strictly hierarchical in nature.

The focus of this study was children's understanding that gender does not change when appearances do, as measured by the Gender Constancy Interview and the A-R Gender Test. However, both of these tasks assess other knowledge beyond the understanding that gender does not change. The Gender Constancy Interview includes a series of questions that assess gender identity and stability, and the A-R Gender Test includes questions that assess the ability to report the phenomenal aspects of a dressed-up child. The scoring criteria for each of these tasks reflects competence with the entire series of questions within each task and may, therefore, reflect children's skill with other than the transformation of gender. Each of the tasks does, however,

have two questions that only assess the transformation of gender. In the Gender Constancy Interview, the two constancy (Stage 3) questions are transformation questions ("If you dressed up in [opposite sex] clothes, would you be a boy or a girl?" and "If you played with [opposite sex] toys would you be a boy or a girl?"). In the A-R Gender Test, the two reality questions ("What is this [the dressed-up child] really and truly, a boy or a girl?") are transformation questions. Children's performance on the transformation questions was examined separately to explore children's understanding that gender does not change.

On the A-R Gender Test, 43% of the three-year-olds, 76% of the four-year-olds, and 100% of the five-year-olds correctly responded to the transformation question. At each age, this is above the level expected by chance alone when tested using the exact binomial test, $p < .025$, $p < .001$, and $p < .001$, respectively. On the Gender Constancy Interview, the variability across the age groups found when all items were included was still evident when only transformation questions were considered: 61% of the three-year-olds, 38% of the four-year-olds, and 56% of the five-year-olds passed. At each age, this is above the proportion expected by chance for the three- and five-year-olds, $p < .001$ and $p < .001$, respectively, but not for four-year-olds, $p < .06$.

Of particular interest was whether the proportion of children passing the A-R Gender Test transformation

questions was significantly greater than that of children passing the Gender Constancy transformation questions, consistent with Prediction 2. This was examined by comparing the percentages of children passing each type of question within each age group. Within the three-year-old group, the percentage of children who passed the Gender Constancy transformation question (61%) was slightly but not significantly higher than the percentage who passed the A-R Gender Test transformation question (43%), $\underline{Z} = 1.35$, $\underline{p} < .09$. However, the trend was reversed in the remaining two age groups. Within the four-year-old group, the percentage of children passing the A-R Gender Test transformation question (77%) was significantly higher than that of children passing the Gender Constancy question (38%), $\underline{Z} = 3.25$, $\underline{p} < .01$. Similarly, 100% of the five-year-olds passed the A-R Gender Test transformation questions whereas only 56% passed the Gender Constancy Interview transformation question, $\underline{Z} = 3.75$, $\underline{p} < .01$. Thus, when other task demands were not included, four- and five-year-old children performed better on the A-R Gender Test transformation questions than on the Gender Constancy Interview transformation questions. Three-year-olds did not differ significantly in their rates of passing each.

Q.2) Is the understanding that gender is unchangeable related to higher levels of sex-typing?

Prediction 1

The prediction that children who passed the A-R Gender Test would have higher sex-typed preference scores was tested with an ANOVA of children's SERLI preference scores. A 2 (Preference type: SRP, SRC) x 3 (Activity type: Adult, Child, Toy) x 2 (A-R Gender Test pass/fail) x 2 (Sex) x 3 (Age group) ANOVA, with type of preference and activity scores as within-subject factors, was conducted. No main effect of having passed the A-R Gender Test was found, nor was there any interaction of performance on the A-R Gender Test with any other factor. Children who passed or failed A-R Gender Test also did not differ in the average scores of sex role knowledge for own- or opposite-sex, nor in the number of sex-typed occupational choices made (see Table 8).

Having passed or failed the A-R Gender Test also did not differentiate children who made sex-typed occupational choices. Chi-square analyses indicated that children who passed or failed the A-R Gender Test did not differ in their choice of sex-typed or non-sex-typed occupational choices, χ^2 (1, N = 87) = 2.62, $p < .11$.

Prediction 2

As predicted, performance on the Gender Constancy interview was unrelated to sex-typed preferences, as indicated by an ANOVA of children's SERLI preference scores. Because few children in the four- and five-year-old group achieved Stage 0 or 1, age groups were combined in a 2

Table 8

Average Preference, Knowledge, and Flexibility Scores, and the Number of Children Sex-Typed Occupational Choices by Children Passing and Failing Appearance-Reality Gender Test

	PASSED		FAILED	
Preference Scores	M	Sd	M	Sd
SRP CHILD	60.97	(13.4)	58.36	(11.6)
SRC CHILD	58.13	(13.1)	56.20	(11.8)
SRP ADULT	60.16	(12.6)	60.30	(11.4)
SRC ADULT	58.67	(10.5)	58.62	(11.8)
SRP TOY	61.71	(12.8)	61.89	(12.5)
SRC TOY	61.65	(13.4)	60.32	(13.7)
Knowledge Scores				
OWN-SEX	94.52	(8.5)	92.86	(10.6)
OPP-SEX	87.42	(12.9)	78.75	(17.8)
Flexibility Scores				
FLEX	5.10	(3.5)	4.71	(3.1)
# Preferring Same-Sex Occupation				
SAME-SEX	20	(65 %)	26	(46 %)

(Preference type: SRP SRC) x 3 (Activity type: Adult, Child, Toy) x 4 (Gender Constancy Stage 0-3) x 2 (Sex) ANOVA, with preference score type and activity type as within-subject factors. No main effect of having passed the Gender Constancy Interview was found, and Gender Constancy Interview Stage did not interact with any other factor.

Summary of the Research Questions

The understanding that gender does not change despite changes in physical appearance, as indicated by performance on the A-R Gender Test, was not reliably demonstrated by the majority of children within an age group until five years of age. Two of the three predictions associated with question 1 were supported. There was a clear developmental trend in the number of children passing the A-R Gender Test, although the number of four- and five-year-olds passing the A-R Gender Test was smaller than had been expected. No relationship was found between the A-R Gender Test and Gender Constancy measure performance, and the A-R Gender Test transformation questions were passed before Gender Constancy transformation questions at four and five years of age. Performance on the A-R Gender Test and the A-R Distinction Test was correlated. When combined, the three tasks showed a developmental trend in the majority of children passing at each age; however, the tasks did not form a strictly hierarchical scale.

The first prediction associated with Question 2 was not

supported. The ability to pass the A-R Gender Test was not related to children's preference for sex-typed activities and occupations. However, the second prediction was supported. As has been found elsewhere, gender constancy was unrelated to sex-typed preferences.

Interestingly, although performance on the A-R Distinction Test and the A-R Gender Test was positively correlated, not all children who passed the A-R Distinction Test passed the A-R Gender Test. At each age level, a large number of children who passed the A-R Distinction Test failed the very similar A-R Gender Test. To determine why children were failing the A-R Gender Test, an analysis of the errors made on the A-R Distinction Test and A-R Gender Test was conducted.

Analyses of the A-R Errors

Differences emerged in the number and types of errors children made on the A-R Distinction Test and the A-R Gender Test, which might offer information about why many children failed the A-R Gender Test but not the A-R Distinction Test. As reported in the preliminary analyses, no difference in the number or type of errors that boys and girls made on the A-R Distinction Test objects was found indicating that, overall, A-R Distinction Test objects were responded to similarly by both sexes. Preliminary analyses also revealed that, as expected, for both the A-R Distinction Test and

the A-R Gender Test phenomenism errors declined with age and did not differ by sex of child or object tested. Thus, the A-R Distinction Test and the A-R Gender Test performance was similar in terms of overall error rates, no sex differences, and the decline of phenomenism errors with age. Where the two tasks did differ, however, was in the rate of realism errors: the number of children making realism errors declined with age for A-R Distinction Test but not for A-R Gender Test. This suggests that different objects yielded different types of realism errors, and was investigated further.

To determine whether the different objects in the A-R Distinction Test and A-R Gender Test elicited different types of errors from boys and girls, Chi-square analyses were conducted. Boys and girls did not differ in the number of realism errors made on the rock/sponge, egg/stone, or the chocolate/magnet. For the A-R Gender Test pictures, however, differences clearly emerged.

Chi-square analyses of the number of boys and girls making realism errors on A-R Gender Test indicated no difference for the picture of the girl dressed as a boy, χ^2 (1, N = 87) = .159, p< .69, but a significant difference in the number of errors for the picture of the boy dressed as a girl. For that picture, boys made more realism errors (43%) than did the girls (21%), χ^2 (1, N = 87) = 4.93, p <.02. The percentage of children making realism errors, by age and

sex, on the picture of the boy as a girl, the girl as a boy, and average A-R Distinction Test objects, is presented in Table 9.

The boys' realism errors on the picture of the boy dressed as a girl become particularly interesting when examined within the five-year-old group alone. As can be seen in Table 9, the proportion of errors five-year-old boys made on the picture of the boy as a girl (46%) was twice that for the picture of the girl as a boy (23%), $\underline{Z} = 1.23$, $\underline{p} < .11$, and far exceeded the average rate of realism errors by five-year-old boys on the A-R Distinction Test in general (8%), $\underline{Z} = 2.01$, $\underline{p} < .02$. In contrast, the five-year-old girls had low rates of errors on the picture of the boy as a girl (17%), on the picture of the girl as a boy (17%), and on A-R Distinction Test objects in general (14%), which did not differ significantly from each other.

To determine whether the five-year-old boys who made realism errors on the picture of the boy dressed as a girl differed from the boys who did not make errors, t-tests of the difference in age in months and Gender Constancy Stage and Chi-square tests of A-R Distinction Test performance were conducted. Results indicated that the two groups of

Table 9

Percent of Children Making Realism Errors on Appearance-Reality Gender Pictures and Average Appearance-Reality Distinction Task by Age Group and Sex

BOYS

(N=44)	Boy-girl		Girl-boy		Mean A-RD
	#	(Percent)	#	(Percent)	(M Percent)
3 Yrs	6	(43 %)	5	(36 %)	(33 %)
4 Yrs	7	(41 %)	6	(35 %)	(25 %)
5 Yrs	6	(46 %)	3	(23 %)	(8 %)

GIRLS

(N=43)	Boy-girl		Girl-boy		Mean A-RD
	#	(Percent)	#	(Percent)	(M Percent)
3 Yrs	2	(14 %)	5	(36 %)	(42 %)
4 Yrs	5	(29 %)	5	(29 %)	(37 %)
5 Yrs	2	(17 %)	2	(17 %)	(14 %)

five-year-old boys did not differ in any respect. Furthermore, the boys who made realism errors were no different than their peers on any of the SERLI subscales. The boys in the five-year-old group who made realism errors on the picture of the boy dressed as a girl were not distinguishable from those who correctly identified the picture on any of the measures in this study.

Realism errors have relevance to the questions of when children understand that gender does not change. If realism errors are non-random, that is, occur more frequently than would be expected by chance alone, they could be considered evidence of the understanding that gender did not change with appearances. In the present sample, when realism errors were included as evidence of children's understanding that gender does not change, 34% of the three-year-olds, 77% of the four-year-olds, and 100% of the five-year-olds indicated that gender had not changed when the child's appearance had. Exact binomial tests indicated that, at each age, this was significantly above the proportion expected by chance ($p < .024$;, $p < .001$, and $p < .001$, respectively). Thus, although children did not reliably report the distinction between apparent and real gender until five years of age, the understanding that the real identity of a dressed-up child was unchanged was reliably reported by the majority of children by four years of age.

Discussion

When do Children Understand that Gender Does Not Change?

The first goal of this study was to develop the Appearance-Reality Gender Test (A-R Gender Test) based upon appearance-reality distinction task procedure to determine when children understand that gender does not change. As predicted, a developmental trend in children's ability to solve the A-R Gender Test emerged. Few three-year-old children passed, almost half the four-year-olds passed, and the majority of five-year-old children passed. Although the ability to pass the A-R Distinction Test, A-R Gender Test, and Gender Constancy Interview emerged in a developmental sequence, the three tasks did not form an hierarchical scale. A-R Gender Test was not passed earlier than the Gender Constancy Interview. Nonetheless, the ability to accurately distinguish real from apparent gender identity was reliably demonstrated by the majority of children at five years of age.

The A-R Gender Test is an identity task and as has been found elsewhere (e.g, Flavell et al., 1983) after three years of age children who failed tended to make realism errors. Three-year-old children made a similar number of realism and phenomenism errors whereas older children made at least twice as many realism than phenomenism errors when responding on the A-R Gender Test. Realism errors have relevance for the question of when children understand that

gender is unchangeable because they occur when information about the real identity of an object interferes with the ability to correctly report the appearance of the object. Therefore, on the A-R Gender Test realism errors reflect children's difficulty ignoring what they know to be the real identity of the dressed-up child when reporting what it looks like.

Although realism errors are definitely a failure to solve the appearance-reality distinction, they indicate knowledge that the real identity of the pictured child has not changed when the appearance has. This is the opposite of a traditional failure of gender constancy, in which children apparently believe that changes in appearance change a child's gender. In the present sample, when realism errors were included as evidence of children's understanding that gender does not change, 34% of the three-year-olds, 77% of the four-year-olds, and 100% of the five-year-olds indicated that gender had not changed when the child's appearance had. Thus, although children did not reliably report the distinction between apparent and real gender until five years of age, the understanding that the real identity of a dressed-up child was unchanged was clearly evident by four years of age, substantially earlier than has been found with the Gender Constancy Interview (Slaby & Frey, 1975).

Relationship Between the A-R Gender Test and Gender Constancy Interview

As predicted, children's performance on the A-R Gender Test was uncorrelated with performance on the Gender Constancy Interview which suggests that, although both tasks assess knowledge of the transformation of gender, procedural differences yielded different performance from children. Contrary to expectations, the A-R Gender Test was not passed more frequently than was Stage 3 (constancy) of the Gender Constancy Interview at each age.

The number of children passing the A-R Gender Test increased steadily with age, whereas the number of children passing Stage 3 of the Gender Constancy Interview varied across the age groups: More three-year-olds than four-year-olds passed Gender Constancy, and only 10% more five-year-olds than three-year-olds passed Gender Constancy. This variability was not explained by within-group subject differences. The three-year-olds who passed gender constancy were not differentiated from their age-mates who failed by age in months, preschool from which they were sampled, or any of the measures in this study. The variability in performance on the Gender Constancy Interview was also not explained by difficulty with the counter-questions. Contrary to what Seigel and Robinson (1987) found, only two children (one four-year-old and one five-

year-old) out of 87 subjects failed an item on the Gender Constancy Interview because they changed their reply when asked the counter-questions.

It is not clear why the three-year-olds in this sample performed so well on the Gender Constancy Interview, although similar performance has been found in other samples (e.g., MacKain, 1986). Given the age-appropriate performance of the three-year-olds in the present sample on the other measures in this study, it seems unlikely that their performance on the Gender Constancy Interview represents sophistication with the concept of unchangeable gender. It is possible that the three-year-olds did not interpret and deal with constancy questions in the same manner as older children. At three years of age, children might have difficulty processing hypothetical questions and might not fully appreciate the implications of them. For example, they might not have "projected" themselves into the future state suggested by the question "if you dressed up...". Instead, the three-year-olds might have been responding to the last phrase of the question "...would you be a boy or a girl?", to which they could be expected to respond correctly, as they would to identity questions. If this is so, the three-year-olds' performance on the Gender Constancy Interview would be the result having responded only to a portion of the question rather than the product of a sophisticated gender concept. However, in the absence of

other information regarding their interpretation of constancy questions this remains speculation. Therefore, it can only be concluded that performance on the A-R Gender Test improved across age groups, whereas performance on the Gender Constancy Interview was inconsistent across age groups.

Of particular interest to this study was children's understanding that gender does not change when appearances do. However, the Gender Constancy Interview includes a series of questions about gender identity and stability, and the A-R Gender Test includes questions about the ability to report the phenomenal aspects of the dressed-up child. Thus, the scoring criteria for each of these tasks reflects competence with the entire series of questions within each task and may reflect children's skill with other than the transformation of gender. The understanding that gender does not change when appearances do is most directly assessed by the transformation questions in the Gender Constancy Interview and the A-R Gender Test (the two constancy questions for the Gender Constancy Interview and the two reality questions for the A-R Gender Test). When transformation questions alone were examined, more children correctly responded to the A-R Gender Test questions than the Gender Constancy questions. For the A-R Gender Test, 43%, 76%, and 100% of the three-, four-, and five-year-olds, respectively correctly responded to the transformation

questions. For the Gender Constancy Interview, the variability across the age groups found when all items were included was still evident when only transformation questions were considered: 61%, 38%, and 56% of the three-, four-, and five-year-olds passed. Thus, when other task demands were not included, children performed better on the A-R Gender Test transformation questions than on the Gender Constancy Interview transformation questions.

Implied in the comparison of the Gender Constancy Interview and A-R Gender Test transformation questions is the assumption that the two procedures assess similar constructs (the transformation of gender) with different procedures. However, the two tasks also differ in other important ways. Specifically, they differ in the use of a concrete vs. hypothetical referent, the assessment of own-sex vs. opposite-sex transformations, and the emphasis upon pretend-real responding.

The use of concrete and hypothetical referents is one way in which the Gender Constancy Interview and the A-R Gender Test differ. The transformation questions in the Gender Constancy Interview are presented verbally, requiring children to solve hypothetical question about a future state. Alternatively, the A-R Gender Test is a visual task: children can see the change enacted in the series of photographs and respond to questions about the child in the last picture. Piaget (1954) suggested that visual

information would impede performance on transformation tasks because of the seductive effect of the perceptual information. However, in a careful assessment of preschool aged children's performance on verbal and visual transformation tasks, Martin and Halverson (1983) found no differences in children's performance as a function of type of presentation. Thus, it is unlikely that differences between performance on the A-R Gender Test and the Gender Constancy Interview were due strictly to the use of concrete versus hypothetical referents.

Another difference between the Gender Constancy Interview and the A-R Gender Test is that the former assesses constancy for the self, whereas the latter assess constancy for another child. It could be argued that entertaining the possibility of changing one's gender might be more anxiety-provoking for children and impede their performance on the task; however, this has not generally been found. When constancy for both the self and another is tested, children perform better on tests of own-sex constancy than of others' (Marcus & Overton, 1978; Gouze & Nadelman, 1980; Martin & Little, 1990) and is generally explained in terms of greater familiarity with one's own sex. If the self-other distinction strongly influences children's performance on transformation questions, in the present sample children would be expected to pass the Gender Constancy Interview transformation questions more frequently

than those of A-R Gender Test, at each age. This was not the case, suggesting that the self-other distinction did not significantly affect children's performance on these two measures.

What, then, accounts for performance differences in the two tasks for assessing the understanding of gender transformations? A major difference between the two measures is that the A-R Gender Test emphasizes real versus pretend responding, whereas the Gender Constancy Interview does not. Preschool children's performance on the Gender Constancy Interview has been found to improve substantially when they are encouraged to give real, rather than pretend, answers. In Martin and Halverson's (1983) study, children were tested with both visual and verbal tests of gender constancy with both traditional instructions and instructions encouraging them to respond in real, rather than a pretend, mode. When children received the traditional instructions, 64% were classified as gender constant. However, when encouraged to give real rather than pretend responses, 95% of the children were classified as gender constant. This effect was not simply due to children's propensity to change their response when reminded not to pretend: children who correctly responded that the cross-dressed child was still the original sex were not influenced to retract their correct judgement when encouraged to respond in a real rather than pretend mode.

Martin and Halverson further found that pretense responding was more prevalent in verbal than visual tasks and concluded that the hypothetical nature of the verbal questions may encourage children to give pretense responses.

In summary, the relationship between performance on the Gender Constancy Interview and the A-R Gender Test can be described as follows. When overall task performance was considered, no correlation between A-R Gender Test and Gender Constancy Interview performance was found. Gender Constancy was passed more frequently than was the A-R Gender Test at each age. However, when transformation questions were considered separately, A-R Gender Test transformation questions were passed more frequently than Gender Constancy transformation questions by four- and five-year-old children. Children more frequently responded correctly to the A-R Gender Test transformation questions despite the focus upon the more difficult other-constancy dimension and despite the presence of strong perceptual cues that could inhibit correct responding. It appears that performance on transformation questions is facilitated by the use of concrete, pictorial presentation and by encouraging children to attend to the object's "real" identity. Because tasks that present transformations verbally may underestimate children's knowledge, the A-R Gender Test may provide an effective method of assessing the early understanding that gender does not change.

Relationship Between Performance on the A-R Distinction Test and the A-R Gender Test

As predicted, performance on the A-R Gender Test and the A-R Distinction Test was positively correlated, suggesting that the same abilities underlie performance on each. However, the two tests did not yield equivalent performance: Within each age group, more children met the criteria for passing the A-R Distinction Test than the A-R Gender Test. This discrepancy is interesting given that the A-R Gender Test was logically derived from appearance-reality distinction procedures. It suggests that children's performance on what is essentially a cognitive task deteriorated when the object in question changed from a non-social object (e.g., rocks and eggs) to one with social relevance (gender).

It is possible that differences in performance on the A-R Distinction Test and the A-R Gender Test were due to different levels of task difficulty. Children had to pass only two of the three A-R Distinction Test items to have been credited with passing the measure, whereas they had to pass both A-R Gender Test picture series to have been credited with passing the A-R Gender Test. The probability of passing the A-R Distinction Test was, therefore, substantially greater than that of passing the A-R Gender Test by chance alone ($p=.1517$ and $.0625$, respectively). However, when performance was examined equalizing the

probability levels, the same results were obtained: when performance on only the first two A-R Distinction Test items was compared with performance on the two items of the A-R Gender Test, a higher percentage of children passed the A-R Distinction Test (51%) than passed the A-R Gender Test (36%), $z = 4.16$, $p < .01$. The difference in chance responding alone does not account for the poorer performance on the A-R Gender Test.

Another manner in which the A-R Distinction Test and the A-R Gender Test differed was in the amount of information presented about the transformation. In the A-R Distinction Test the change was described verbally. In the A-R Gender Test the photographs visually depicted the transformation. Performance on such tasks has been found to be improved by providing children with visual information (Keil, 1979) or verbal information (MacKain, 1986) regarding the transformation which, in the present study, should have resulted in better performance on the A-R Gender Test than the A-R Distinction Test. This was not the case. Thus, it appears that content differences and not procedural differences may have been responsible for the poorer performance on the A-R Gender Test than on the A-R Distinction Test. Further consideration of the effect of gender on appearance-reality distinction task performance is included in the discussion of the sex differences in A-R Gender Test errors.

The Unchangeability of Gender and the Development of Sex-Typing

The second major goal of this study was to re-address Kohlberg's (1966) proposal that the understanding that gender does not change is related to the acquisition of sex-typing. If Kohlberg's proposal is correct, children who passed the A-R Gender Test would be expected to have higher indices of sex-typing on the SERLI, and this prediction was not supported. Children who passed and failed the A-R Gender Test did not differ in their preference for sex-typed adult roles, child activities, or toys. Children who passed and failed the A-R Gender Test also did not differ in their preference for sex-typed occupational choices. Furthermore, when children's realism errors were included as evidence of understanding the unchangeability of gender, the same results were obtained. Children who understood that gender does not change when appearances do did not differ on any of the sex-typing indices used in this study.

The results of this study, using a different procedure for assessing the understanding of gender, are similar to those found in studies using the Gender Constancy Interview (Coker, 1984; Carter & Levy, 1985, 1987; Downs & Langlois, 1988; Martin & Little, 1990) and suggest that Kohlberg's conceptualization of the process by which children internalize sex-typed information into prescriptions for their own behaviour is inaccurate. In the present study,

performance on neither the A-R Gender Test nor the Gender Constancy Interview was related to sex-typed preferences.

One reason that no relationship emerged between understanding the unchangeability of gender and sex-typing could be that the use of preferences as the criterion measure did not adequately capture children's sex-typing. With preschool-aged children, sex-typing is usually assessed by the degree to which they demonstrate preferences for sex-typed toys, objects, activities, playmates, or adult occupations. Children are asked to indicate which items they prefer from a sample of pictures or objects and the proportion of sex-typed choices the children make is considered indicative of their degree of sex-typing. Of course, sex-typing extends to areas of behaviour other than play preferences. It has been examined in terms of sex differences in children's play styles (Fagot, 1984; Smetana & Letourneau, 1984), personal and social attributes (Fagot, 1981), and proximity to adults (Jacklin & Maccoby, 1978). However, in terms of Kohlberg's model, sex-typed preferences more accurately reflect the critical element in Kohlberg's proposal, that once children have determined that gender is unchangeable they begin to value sex-appropriateness. Preferences may be a more accurate assessment of valuation than behaviour because many factors other than sex-appropriateness influence children's behaviour. Actual play behaviour is shaped by such things as availability of toys,

sex of available playmates, and reinforcement by adults or peers. For example, a girl with sex-typed toy preferences might decline to play with dolls if her preschool teacher strongly discourages sex-typed play. Or, a girl may resist her preference to play in the housekeeping centre if the boys are playing war games in that area. Preferences are somewhat removed from these immediate influences and may better reflect children's valuation of sex-types. Furthermore, as Kohlberg suggested, preferences may influence the degree to which children attend to and accumulate information about sex roles, whereas environmental influences probably primarily affect the situations in which sex-typed preferences are expressed. Therefore, although preferences may not be the most comprehensive index of sex-typing, they sufficiently capture the essence of Kohlberg's (1966) proposed model.

The failure to find a relationship between the understanding that gender does not change and the development of sex-typing, even when the understanding is demonstrated at four years of age, suggests that the development of sex-typing is independent of the concept of gender. Gender labels may be sufficient for the organization of information by children and the subsequent prescription for behaviour (Fagot, 1985). In a study of the contribution of cognitive, social, and affective variables to the development of sex-typing, Serbin, Powlishta, and

Gulko (1992) found that social factors, such as parents' sex-typing, was more closely related to children's sex-typed preferences than were cognitive variables. Thus, understanding the cognitive development of 'constancy' may be tangential to understanding the development of sex-typing. The three-year-old children in the present sample, many of whom did not pass the cognitive measures of constancy or appearance-reality distinction, already possessed an impressive fund of sex-role knowledge and sex-typed preferences, suggesting that the affective components of sex-typing are influential before cognitive competence with the concept of gender is evident.

Sex Differences in A-R Gender Test Realism Errors

An examination of the types of errors that children made revealed that boys made substantially more realism errors on the picture of the boy dressed as a girl than on the picture of the girl dressed as a boy, despite the absence of any sex differences in phenomenism errors on both trials of the A-R Gender Test or all three A-R Distinction Test objects. Furthermore, boys made more same-sex realism errors than did the girls on either picture or than they did on the A-R Distinction Test items combined. This effect was most striking for the five-year-old boys for whom the rate of same-sex realism errors was twice that for their female age-mates and significantly higher than their overall rate

of realism errors on the A-R Distinction Test items.

This finding was unexpected but compelling: what would lead the boys to make more realism errors on the picture of the boy dressed as a girl in the A-R Gender Test? It is unlikely that there were differences in the levels of cognitive development between boys and the girls in the present sample given that the boys did not differ from the girls in their performance on other A-R Distinction Test items or on any other measure in this study. It is also unlikely that the boys' realism errors represented deficits in skills or abilities that were expressed singularly on the A-R Gender Test. The difference is also not explained by any feature specific to the picture of the cross-dressed boy, because both boys and girls would have then been affected similarly.

The knowledge that cross-dressing in general is inappropriate also does not explain the discrepancy between boys' and girls' realism errors, because boys and girls did not differ in their overall levels of sex-role knowledge and girls did not make a similar number of errors on their same-sex picture. It is possible that male cross-dressing in general is considered more socially inappropriate than female cross-dressing in general, and that the stronger taboo resulted in a reluctance to report that the boy looked like a girl. However, if a general taboo against male cross-dressing was affecting children's performance, again,

both the boys and the girls should have made similar errors on the picture of a boy dressed as a girl. If a taboo against male cross-dressing was responsible for the realism errors on the picture of a boy dressed as a girl, it was a taboo that was evident only to the boys.

Given the boys' competence with the three A-R Distinction Test items (the rock, chocolate, egg) and their ability to identify the sex of the other-sex cross-dressed child, it is striking that five-year-old boys were reluctant to report that the boy dressed as a girl looked like a girl. It suggests that their errors occurred for affective, rather than cognitive reasons. One such reason could be that cross-sex behaviour has more negative consequences for boys than girls.

Parents give more non-verbal support to toddlers engaged in sex-typed play than to those engaged in cross-sex play (Fagot, 1977) and respond differently to boys' and girls' activities (Fagot, 1984). Fathers in particular give negative responses to boys engaged in cross-sex activities but reward sex-typed play in both boys and girls (Caldera, Huston, and O'Brien, 1989). Fagot (1977) found that preschool aged boys received greater amounts of criticism from teachers when involved in dress-up games than girls did when involved in cross-sex activities.

Boys also receive negative feedback for sex-inappropriate play from peers. Preschool aged boys who

engage in cross-sex activities are more frequently criticized by peers than are girls, are avoided as playmates, and are three times more likely to play alone (Fagot, 1977; Carter & McClosky, 1984; Stoddart & Turiel, 1985). Lamb and Roopnarine (1979) found that boys appear more concerned than girls about supporting and reminding others of sex-appropriate behaviour: as well as giving more negative feedback for cross-sex play, boys positively reinforce sex-appropriate behaviour for both boys and girls. Lamb and Roopnarine further found that while negative feedback served to decrease children's sex-inappropriate behaviour it had no effect upon sex-appropriate behaviour, suggesting that negative feedback does not appear to alter behaviour unless it is sex-inappropriate. Thus, one possible reason why the boys in the present study made realism errors on the picture of the boy dressed as a girl is that they were discomforted by the sex-inappropriate activity, understanding that negative consequences usually follow such behaviour.

Although boys' experience of negative consequences may explain the affective response to and avoidance of cross-sex behaviour, it does not explain why adults and peers have more rigid sex-role expectations of boys in the first place. In fact, there is evidence to suggest that young children do not believe that sex role transgressions are intrinsically wrong. Carter and McClosky (1984) found that the majority

of five-year-old children in their sample said that they did not believe that cross-sex behaviour was wrong in the same sense that a moral transgression, such as lying, was. Similarly, Smetana (1986) found that at three and four years of age children believed sex-role behaviour was a matter of personal choice and social convention, and rarely indicated that serious consequences, such as loss of gender or fear of punishment, were reasons to avoid cross-sex behaviour. Thus, although punishment or negative consequences may influence boys to avoid cross-sex behaviour, why they are motivated to continue such socialization practices is not clear.

One reason why more rigid sex-role expectations exist for boys than for girls could be that adults respond to perceived status differences in male and female sex-roles. If adults believe that masculine activities are more prestigious or important than feminine activities, girls may be permitted to engage in cross-sex behaviour because it increases their status, whereas boys may be discouraged, believing that they are compromising their status. Young children may internalize perceived status differences and adhere to their sex roles as a means of maintaining status. In a study of preschool-aged children's emulation of same and other-sex models, Bussey and Bandura (1984) found that, while children of both sexes prefer to imitate same-sex models, children will imitate other-sex models more if the

other-sex model is perceived as having power. This effect was slightly stronger for boys than for girls, which led Bussey and Bandura to suggest that boys were more attracted to the powerful model because they believed that the higher status was consistent with their sex role, regardless of the actual sex of the model. While this raises the possibility that boys believe that power is sex-appropriate, it does not explain the girls' performance. If children imitate the models that reflect sex-appropriate activity, girls should have been attracted to the less-powerful model, regardless of the actual sex of the model. This was not the case. Girls were attracted to female models when all other factors were held constant, to female models when they were perceived as powerful, and to male models when they were perceived as powerful. Thus, it appears that having power is perceived by both sexes as desirable, and it is not clear whether the children shared the perception that being more powerful is a male attribute or whether the boys simply more rigidly adhered to their sex role than the girls did.

A third possible explanation of the source of rigid sex-role expectations for boys is that boys themselves feel less secure of their gender identity and require greater sex-role adherence. Boys' rigid adherence to sex-appropriateness resembles the process described by Dipboye (1977) of identity affirmation. Dipboye proposed that identity affirmation is undertaken when individuals

experience a sense of insecurity about group membership, which is considered an anxiety-provoking experience. Identity affirming behaviours are a form of display behaviours, which clearly announce to observers that the person in question is a member of a particular group and serve to reduce anxiety. The concept of identity affirmation was specifically proposed in terms of social group membership but may be informative in considering the issue of gender identity. Gender is an important social category and children quickly learn that everyone is a member of one sex or the other. If a child has a sense of ambiguity or conflict regarding his or her gender group membership, it might provoke anxiety and result in identity-affirming behaviours such as the boys' stereotype rigidity described above.

The possibility that boys experience greater difficulty and conflict establishing their gender identity has been proposed by two theorists. Freud (1925) suggested that there are fundamental differences in how boys and girls respond to learning the biological differences between the sexes. Freud proposed that children respond to the observation that males have a penis and females have a vulva by assuming that females have had their penises removed. Boys were thought to experience intense anxiety regarding the possibility that their penis could be lost, and that anxiety was proposed to play an instrumental role in the

boys' identification with their fathers as an attempt to assuage their fears of losing their penis. Thus, until the resolution of this conflict at about six years of age boys were proposed to experience anxiety regarding the tenuousness of their gender.

Recently, theorists interested in mother-child attachment have examined other possible reasons for sex differences in the establishment of gender identity. Lerner (1974) proposed that boys have more difficulty consolidating their gender identity in the preschool years because they face the extra task of differentiating their gender from that of their primary attachment figure, usually their mothers. Lerner has suggested that both boys and girls similarly experience early individuation from the mother, representing the end of the symbiotic mother-infant relationship. However, boys face a second "separation" when they learn that two sexes exist, by about age two and a half, and then recognize that they and their mothers do not share the same gender. For girls, the process of gender identity is thought to be facilitated by being able to identify with their mothers and maintaining their primary attachment figures as models for gender as well. For boys, however, Lerner believed that the relative uninvolvement of fathers in infants' and toddlers' lives left young boys with the initial process of defining gender by enumerating the ways in which they are different from their mothers. She

believed that, as boys grow older and learn more information about the features associated with being male, they use sex-role information to define their sex. According to Lerner, however, boys' earliest understanding of their gender identity is by exclusion (how they are unlike mother) rather than inclusion (what boys are like), until late in the third year when they begin relying upon their growing awareness of sex-role information from their larger culture. Thus, she proposed, boys may rely more upon external features (activities, appearance, etc.) to define their gender, and have greater investment than girls do in engaging in sex-appropriate behaviour.

While each of these three possible explanations may account for why boys avoid cross-sex activities more than do girls, they do not explain why this phenomenon is stronger for older children than younger ones. If perceived negative consequences for cross-sex behaviour, perceived status differences, or greater need for sex-role adherence cause boys to have lower tolerance for cross-sex behaviour, one would expect it to be also evident in the performance of the boys three and four years of age, or, should begin to emerge with gender labelling or the increase in sex-role knowledge. However, not until about five years of age did the boys in the present study make a substantial number of realism errors on the picture of a boy dressed as a girl relative to their overall performance. Although it seems counter-

intuitive that the five-year-olds would appear less tolerant than the three-year-olds, it is somewhat more comprehensible when the effect of the emergence of Level 2 thinking is considered.

As described earlier, with the advent of Level 2 thought children begin to understand that people can have different mental representations of objects, and the potential for misidentification (e.g., that a boy could be mistaken for a girl) is realized. The potential for misidentification would directly threaten attempts at identity-affirmation. The realization that one could be misidentified would, therefore, be most distressing for children attempting to affirm their gender identity through rigid adherence to sex-appropriate behaviour and heighten intolerance of sex-role transgressions. If these premises are correct, it could be anxiety or ambiguity about gender identity combined with the understanding of the appearance-reality distinction (and, therefore, the potential for misidentification) that caused the older boys in the present study to make realism errors. The affect associated with the possibility that the dressed-up child could be misidentified may have caused them to make realism errors in an attempt to emphasize the child's real gender.

Of course, the data from the present study do not address these speculations. In order to address the possibility that children's responses on the A-R Gender Test

reflected differences in the perception of negative consequences for cross-sex behaviour, one would have to consider the relationship between perceived differences in status of masculine and feminine behaviours, the stability of their gender identity, the relationship between affective components of gender identity, the strength or solidity of gender identity, the ability to distinguish appearance from reality, and adherence to sex stereotypes. In the absence of such information, these speculations serve only to underscore the results obtained from the comparison of the A-R Gender Test and the A-R Distinction Test performance: gender is an affect-laden concept for young children. Assessing children's understanding of gender from a strictly cognitive perspective may be insufficient to capture children's understanding of gender, particularly if their affective response interferes with their performance, as it may have on the A-R Gender Test.

Conclusions

The acquisition of sex-typing is such a complex and robust phenomenon that the belief that gender is an unstable, changeable characteristic apparently does not impede the development of children's preference for sex-typed activities. The children in the present study demonstrated an impressive amount of knowledge about sex roles and strong preferences for sex-appropriate activities before they

indicated the belief that gender is a permanent characteristic. Interestingly, three-year-old children's preference for sex-appropriate activities was evident despite their relatively low level of sex role knowledge, suggesting that preference for sex-typed activities develops earlier than children's content knowledge regarding gender. This is contrary to Kohlberg's (1966) conceptualization of cognitive developmental factors as precedent in the acquisition of sex-typed preferences.

What, then, accounts for the surprisingly consistent and reliable manner in which children develop sex-typing? As Fagot and Leinbach (1985) have suggested, gender labelling, at about two and a half years of age, may be sufficient for children to begin acquiring sex-typed preferences. The advent of labelling, with its inherent understanding that two sexes exist and that certain activities are associated with each, may be sufficient to cause children to organize information in a sex-typed fashion, after which modelling and reinforcement may explain the early manifestations of sex-typed behaviour. Clearly, the development of sex-typed preferences is not dependant upon the understanding that gender does not change.

An important contribution of the present study was the reframing of questions about the understanding that gender does not change into an appearance-reality distinction task, thus avoiding some of the methodological problems associated

with the Gender Constancy Interview (Slaby & Frey, 1975). Using the A-R Gender Test, interesting results were obtained. As expected, children generally passed the A-R Gender Test earlier than they passed the Gender Constancy Interview questions. However, children's performance on the A-R Gender Test was not as good as it was on standard A-R Distinction Test, suggesting that, although they could differentiate appearance from reality for other objects, when the concept in question involved gender their performance on an otherwise standard cognitive task deteriorated.

The possibility that affective components were interfering with children's performance was suggested by the performance of the five-year-old group, who were otherwise very competent making the appearance-reality distinction but appeared reluctant to report that a boy dressed as a girl looked like a girl. The three-year-olds' performance in the present study further suggested that strong affective components of sex-typing, indicated by the preference for sex-appropriateness, emerges early in development and may precede the acquisition of sex role knowledge. The role of affective components has been identified as an important, independent factor in the development of sex-typing (Serbin, Powlishta, & Gulko, 1991), but remains to be fully explored.

Although the A-R Gender Test provided important information about children's understanding of the

unchangeability of gender, this study would have been improved by the inclusion of more trials in the A-R Gender Test, using several picture-series of boys and girls cross-dressing, for two reasons. First, as discussed earlier, in order to pass the A-R Gender Test, children had to pass both questions of both picture-series, yielding a stringent criterion. The inclusion of more trials might have yielded results more similar to that on the A-R Distinction Test, enabling more exact comparisons. A second reason why more trials would have been beneficial is that children may have been surprised by the initial presentation of cross-dressed children and their responses may have been affected by the unfamiliarity. The affective response of many children to the picture of a boy dressed as a girl may have prevented them from correctly responding on the A-R Gender Test, initially. If so, repeated trials might have helped them overcome their hesitancy and revealed performance on the A-R Gender Test slightly more comparable to that on the A-R Distinction Test.

Future studies would benefit from examining an important area not addressed by this study, namely, the effect of children's exposure to sex role information upon the development of gender concepts. Specifically, children may use different sources of sex role information differently when determining what information is appropriate for themselves. During testing in the present study, it was

twice observed that children indicated sex-role knowledge that was directly contradicted by their own experience. For example, one four-year-old girl remarked that being a doctor was for boys, but when asked what she wanted to be when she grew up, she said "a doctor, like my Mom." The sex role models that children use to build their gender concepts may have differential weight when determining what is appropriate for boys and girls, and what is appropriate for themselves. In building a model of how children learn and enact their beliefs about sex-appropriateness, the role of experience needs to be considered.

Finally, the results of this study underscore the conceptualization of sex-typing as a "category under construction from birth" (Fagot & Leinbach, 1985). The interplay between children's cognitive-development, social experience, and affectional ties to sex-appropriateness must be considered when exploring the development of sex-typing. Through the continued study of preschool children at different points in their development, the relationship between when children have sufficient information to recognize gender as a prescriptive social category, what motivates children to be sex-appropriate, and what supports their continued sex-typing throughout development may be explained. Clearly, the meaning and function of preschool children's gender concepts have yet to be fully understood.

References

Aboud, F. (1984). Social and cognitive basis of ethnic identity constancy. Journal of Genetic Psychology, 145 (2), 217-230.

Astington, J.W., & Gopnik, A. (1988). Children's understanding of representational change and its relationship to the understanding of false beliefs and the appearance-reality distinction. Child Development, 59 (1), 26 - 37

Bem, S. L. (1981). Gender schema theory: A cognitive account of sex-typing. Psychological Review, 88 (4), 354-364.

Brown, S.R., & Pipp, S. (1991). The role of the Appearance-Reality Distinction and the genital basis of gender on the development of gender constancy. Paper presented at Society for Research in Child Development, Seattle, Wash. April, 1991.

Bussey, K., & Bandura, A. (1984). Influence of gender constancy and social power on sex-linked modelling. Journal of Personality and Social Psychology, 47 (6), 1292-1302.

Caldera, Y.M., Huston, A.C., & O'Brien, M. (1989). Social interactions and play patterns of parents of toddlers with feminine, masculine, and neutral toys. Child Development, 60, 70-89.

Carter, D.B., & Levy, G.D. (1985). Cognitive aspects of early sex-role development: The influence of gender schema on preschooler's memories and preferences for sex-typed toys and activities. Child Development, 59, (3), 782-792.

Carter, B., Levy, G., & Cappabianca, J. (1985). Stereotyped knowledge, flexibility, and gender constancy: Applications of gender schema theory to sex-typing by preschoolers. Paper presented at the Society for Research in Child Development, Toronto, Canada, April 1985.

Carter, D.B., & McClosky, L.A. (1984). Peers and the maintenance of sex-typed behaviour: The development of children's concepts of cross--gender behaviour in their peers. Social Cognition, 2, 294-314.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, 20 (1), 37-46.

Coker, D.R. (1984). The relationship among gender concepts and cognitive maturity. Sex Roles, 10 (1/2), 19-31.

DeLucia, L.A. (1963). The toy preference test: A measure of sex role identification. Child Development, 34, 107-117.

DeVries, R. (1969). Constancy of generic identity in the years 3-6. Monographs of the Society for Research in Child Development, Serial # 127, 34.

Dipboye, R.L. (1977). Alternative approaches to de-individuation. Psychological Bulletin, 84 (6), 1057-1075.

Downs, C.A., & Langlois, .H. (1988). Sex-typing: Construct and measurement issues. Sex Roles, 18 (1/2), 87-100.

Edelbrock, C., & Sugawara, A.I. (1978). Acquisition of sex-typed preferences in preschool-aged children. Developmental Psychology, 14 (6), 614-623.

Emmerich, W., & Shepard, K. (1984). Cognitive factors in the development of sex-typed preferences. Sex Roles, 11 (1/2), 997-1007.

Emmerich, W., Goldman, K.S., Kirsch, B., & Sharabany, R. (1977). Evidence for a transitional phase in the development of gender constancy. Child Development, 48, 930-936.

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

Fagot, B.I. (1977). Consequences of moderate cross-gender behaviour in preschool children. Child Development, 48, 902-907.

Fagot, B.I. (1981). Continuity and change in play styles as a function of sex of the child. International Journal of Behavioral Development, 4, 37-43.

Fagot, B.I. (1984). The child's expectations of differences in adult male and female interactions. Sex Roles, 11 (7/8), 593-600.

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

Fagot, B.I., & Leinbach, M.D. (1985). Gender identity: Some new thoughts on an old concept. Journal of the American Academy of Child Psychiatry, 24 (6), 684-688.

Flavell, J.H., (1988). The development of children's knowledge about the mind: From cognitive connections to mental representations. In J. Astington, P. Harris, & D. Olsen (Eds.) Developing Theories of Mind. New York: Cambridge University Press.

Flavell, J.H., Flavell, E.R., & Green, F.L. (1983). The development of the appearance-reality distinction. Cognitive Psychology, 15, 95-20.

Flavell, J., Green, F.L., & Flavell, E. (1986). Development of Knowledge about the Appearance-Reality Distinction. Monographs of the Society for Research in Child Development, Serial #212, 51 (1).

Freud, S. (1925). Some psychological consequences of the anatomical differences between the sexes. In J. Strachey (Ed.) Collected Papers, 5. London: Hogarth Press.

Gelman, S.A., & Coley, J.D. (1990). The importance of knowing a Dodo is a bird. Developmental Psychology, 26 (5), 796-804.

Gelman, S.A., Collman, P., & Maccoby, E.E. (1986). Inferring properties from categories versus inferring categories from properties: The case of gender. Child Development, 57, 396-404.

Gelman, S.A. & Markman, E.M. (1987). Young children's inductions from natural kinds: The role of categories and appearances. Child Development, 58, 1532-1541.

Gouze, K.R., & Nadelman, L. (1980). Constancy for self and others in children 3 - 7. Child Development, 51, 275-278.

Guardo, C.J., & Bohan, J.B., (1971). Development of a sense of self-identity in children. Child Development, 42, 1909-1921.

Huston, A. (1983). Sex typing. In P. Mussen (Ed.) Handbook of Child Psychology, (4th Ed.). New York: John Wiley.

Jacklin, C.N., & Maccoby, E.E. (1987). Social behaviour at 33 months in same-sex and mixed-sex dyads. Child Development, 49, 557-569.

Johnson, C.N., (1990). If you had my brain where would I be? Children's understanding of the brain and identity. Child Development, 61, 962-972.

Keil, F.C. (1989). Semantic and Conceptual Development: An Ontological Perspective. Cambridge Mass: Cambridge University Press.

Kohlberg, L. (1966). A cognitive developmental analysis of children's sex role concepts and attitudes. In E.E. Maccoby (Ed), Development of Sex Differences. California: Stanford University Press.

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

Lamb, M.E., & Roopnarine, J.L. (1979). Peer influences upon sex role development in preschoolers. Child Development, 50, 1219-1222.

Lerner, H. (1974). Early origins of envy and devaluation of women: Implications for sex-role stereotypes. Bulletin of the Menninger Clinic, 38 (6), 538-553.

Levy, G., Haaf, R.A., & Sommer, K.L. Infants' detection of correlated features among social stimuli: A precursor to stereotyping? Paper presented at Society for Research in Child Development, Seattle Washington, April, 1991.

Mackain, S.J. (1986). Gender Constancy: A realistic approach. Paper presented at Society for Research in Child Development, April, 1986.

Marcus, D.E., & Overton, W.F. (1978). The development of gender constancy and sex role preferences. Child Development, 49, 434-444.

Martin, C.E. (1989). Children's use of gender-related information in making social judgements. Developmental Psychology, 25 (1), 80-88.

Martin, C.E., & Halverson, C. (1983). Gender Constancy an: A methodological analysis. Sex Roles, 9 (7), 775-790.

Martin, C., & Little, J. (1990). The relation of gender understanding to children's sex-typed preferences and stereotypes. Child Development, 61, 1427-1439.

Moses, L.J., & Flavell, J.H. (1990). Inferring false beliefs from actions and reactions. Child Development, 61, 929-945.

Nie, N., & Hull, H. (1975). Statistical Packages for the Social Sciences. McGraw-Hill.

O'Keefe, E., & Shibley-Hyde, J. (1983). The development of occupational sex role stereotypes: The effects of gender stability and age. Sex Roles, 9 (4), 481-492.

Piaget, J. (1954). The Construction of Reality in the Child. New York: Basic.

Piaget, J. (1968). On the Development of Memory and Identity in Children. Clark University Press with Barre Publishers:New York.

Poulin-Dubois, D., Serbin, L., Kenyon, B., & Derbyshire, A. (In Press). Infants' intermodal knowledge about gender. Developmental Psychology.

Serbin, L.A., & Sprafkin, C. (1986). The salience of gender and the process of sex typing in three- to seven-year-old children. Child Development, 57, 188-1199.

Serbin, L.A., Powlishta, K.K., & Gulko, J. (1993). The Development of Sex-Typing in Middle Childhood. Monographs of Society for Research in Child Development, (Serial no. 232).

Siegal, M., & Beattie, K. (1991). Where to look first for children's knowledge of false beliefs. Cognition, 38 (1), 171-182.

Siegal, M., & Robinson, J. (1987). Order effects in children's gender constancy responses. Developmental Psychology, 23 (2), 283-286.

Siegler, R.S. (1991). Children's Thinking: Second Edition. New Jersey: Prentice-Hall.

Slaby, R.G., & Frey, K.S. (1975). Development of gender constancy and selective attention to same sex models. Child Development, 46 (4), 849-856.

Smetana, J. (1986). Preschool children's conception of sex role transgressions. Child Development, 57, 862-871.

Smetana, J.G., & Letourneau, K.J. (1984). Development of gender constancy and children's sex-typed free play behaviour. Developmental Psychology, 20 (4), 691-696.

Stoddart, T., & Turiel, L. (1985). Children's conception of cross-gender activity. Child Development, 56, 1242-1252.

Tabachnik, B., & Fidell, L. (1989). Using Multivariate Statistics, Second Edition. Harper-Collins Publisher: New York.

Taylor, D.B., & Carter, G.D. (1987). The association between children's gender understanding, sex role knowledge, and sex role preferences. Child Study Journal, 17 (3), 185-196.

Trautner, H. (1985). The significance of the Appearance-Reality Distinction for the development of Gender Constancy. Paper presented at the Biannual conference of the Society for Research in Child Development, Toronto, Canada, April 1985.

Warren, O., & Von Bargen, D. (1981). Asynchronous development of gender understanding in preschool children. Child Development, 52, 1020-1027.

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

Weisner, T.S., & Wilson-Mitchell, J.E. (1990). Nonconventional family lifestyles and sex-typing in 6-year-olds. Child development, 61, 1915-1933.

Welch, R.L., Huston-Stein, A., Wright, J.C., & Plehal, R. (1979). Subtle sex-role cues in children's commercials. Journal of Communication, 29 (3), 202-209.

White, R. (1959). Motivation reconsidered: The concept of competence. Psychological Review, 66, 297-333.

APPENDIX A

A-R Gender Test Procedure

SUBJ# ____

Appearance-Reality Gender

	REPLY	
BOY DRESSED AS A GIRL:		IDBOYG
"What does this look like?"	_____	____
"Now, when you look at this with your eyes right now, does it look like a GIRL or does it look like a BOY?"	_____	APP B as G ____
"What is it, really and truly? Is it really and truly a BOY or is it really and truly a GIRL?"	_____	REAL B as G ____
Perceived Similarity Q: "Are these pictures of the same person, or of two different people?"	_____	SIMBG ____
GIRL DRESSED AS A BOY:		IDGIRLB
"What does this look like?"	_____	____
"Now, when you look at this with your eyes right now, does it look like a BOY or does it look like a GIRL?"	_____	APP G as B ____
"What is it, really and truly? Is it really and truly a GIRL or is it really and truly a BOY?"	_____	REAL G as B ____
Perceived Similarity Q: "Are these pictures of the same person, or of two different people?"	_____	SIMGIRLB ____

APPENDIX B

Gender Constancy Interview Procedure

Identity / Labelling		REPLY	CODE
1) (BOY)	"Is this a girl or a boy?"	_____	____
	"Is it a girl?"	_____	____
2) (GIRL)	"Is this a girl or a boy?"	_____	____
	"Is is a boy?"	_____	____
3) (MAN)	"Is this a lady or a man?"	_____	____
	"Is it a lady?"	_____	____
4) (Woman)	"Is this a lady or a man?"	_____	____
	"Is it a man?"	_____	____
5)	"Are you a girl or a boy?"	_____	____
6)	"Are you a (opposite sex?)"	_____	____
Stability			
7)	"When you were a little baby, were you a little girl or a little boy?"	_____	____
	"were you a (opposite sex)?"	_____	____
8)	"When you grow up will you be a mommy or a daddy?"	_____	____
	"will you be a (opposite)?"	_____	____
Constancy			
9)	"If you wore (opposite sex) clothes, would you be a girl or a boy?"	_____	____
10)	"Would you be a [opposite sex]?"	_____	____
11)	"If you played (opposite sex) games, would you be a girl or a boy?"	_____	____
12)	"Would you be a (opposite sex)?"	_____	____
13)	"Could you be a (opposite sex child) if you wanted to be?"	_____	____

APPENDIX C

A-R Distinction Procedure

Appearance-Reality Distinction

	REPLY	CODE IDROCK
1) Rock / Sponge		
"What does this look like?"	_____	____
"When you look at this with your eyes right now, does it look like a ROCK or does it look like a SPONGE?"	_____	APP-ROCK ____
Reality		
"What is it, really and truly? Is it really and truly a SPONGE, or is it really and truly a ROCK?"	_____	REAL-ROCK ____
2) Chocolate / Magnet		IDCHOC
"What does this look like?"	_____	____
"Now, when you look at this With your eyes right now, does it look like a CHOCOLATE or does it look like a MAGNET?"	_____	APP-CHOC ____
"What is it really and truly? Is it really and truly a MAGNET or really and truly a CHOCOLATE?"	_____	REAL-CHOC ____
3) Egg / Stone		IDEGG
"What does this look like?"	_____	____
"Now, when you look at this with your eyes right now does it look like an EGG or does it look like a STONE?"	_____	APP-EGG ____
"What is it, really and truly? Is it really and truly a STONE or is it really and truly an EGG?"	_____	REAL-EGG ____

APPENDIX D
SERLI Pictures and Procedure

SERLI PREF.

1) Child Figs.

Hammering Ironing Sewing

Cooking Car Play Digging Sweeping

Baseball Fighting Dishes

"If you could do any one of these things, what would you like to do best?"

---

2) Adult Figs.

Feeding Baby Pouring Sawing

Police Teacher Brushing Hair Fire Fighter

Soldier Peeling Apples Doctor

"Which one of these things would you like to do or be when you grow up?"

---

3) Toy Choices

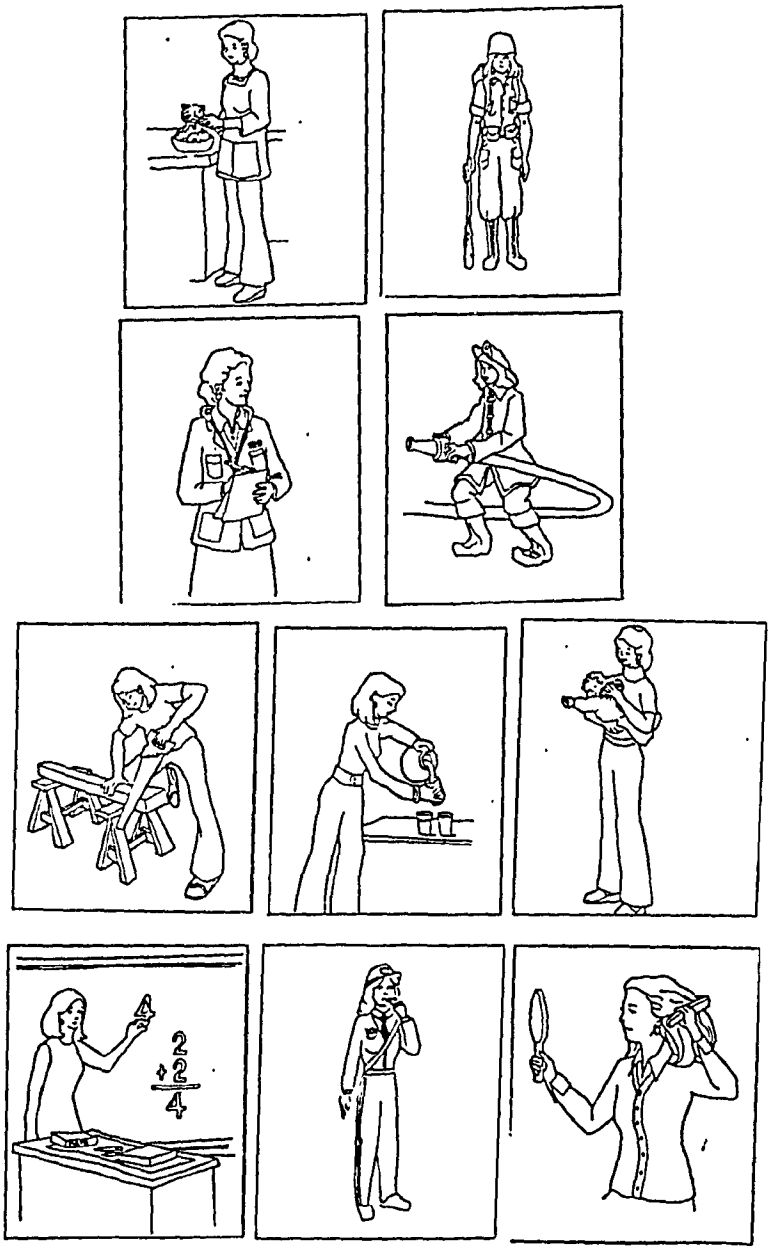
Dress-up M Colouring Dress-up F

Bldg. Blck Skip. Rope Wagon Doll

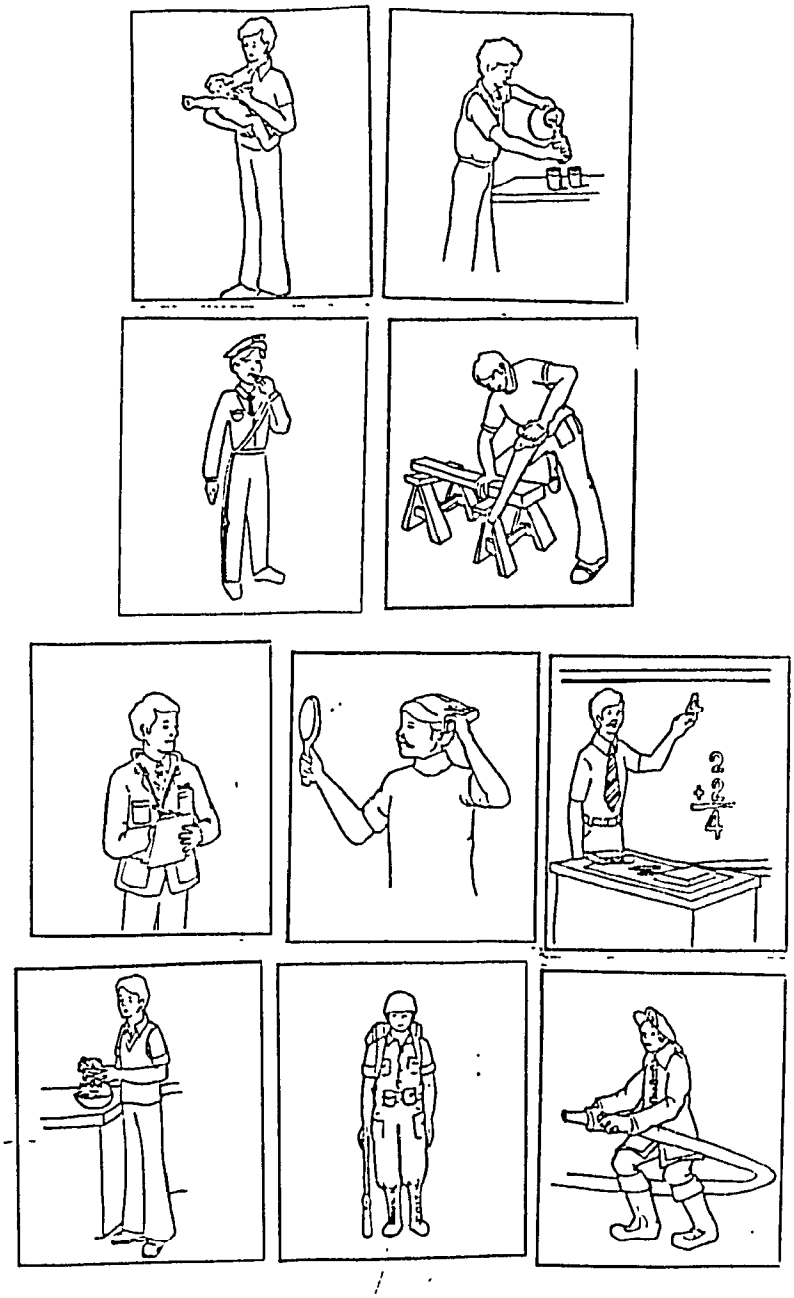
Tea Set Toy Robot Tool Set

"If you could play with any one of these things, which would you like to play with best?"

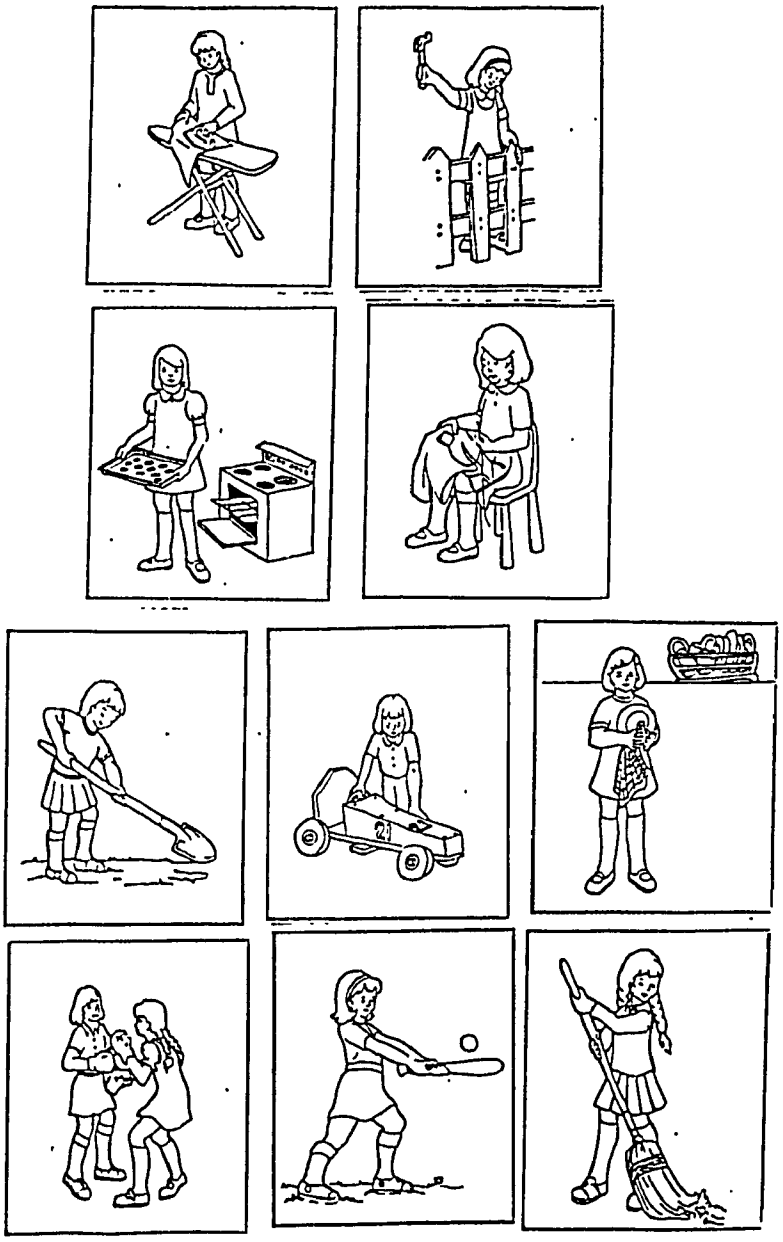
APPENIX D (continued)



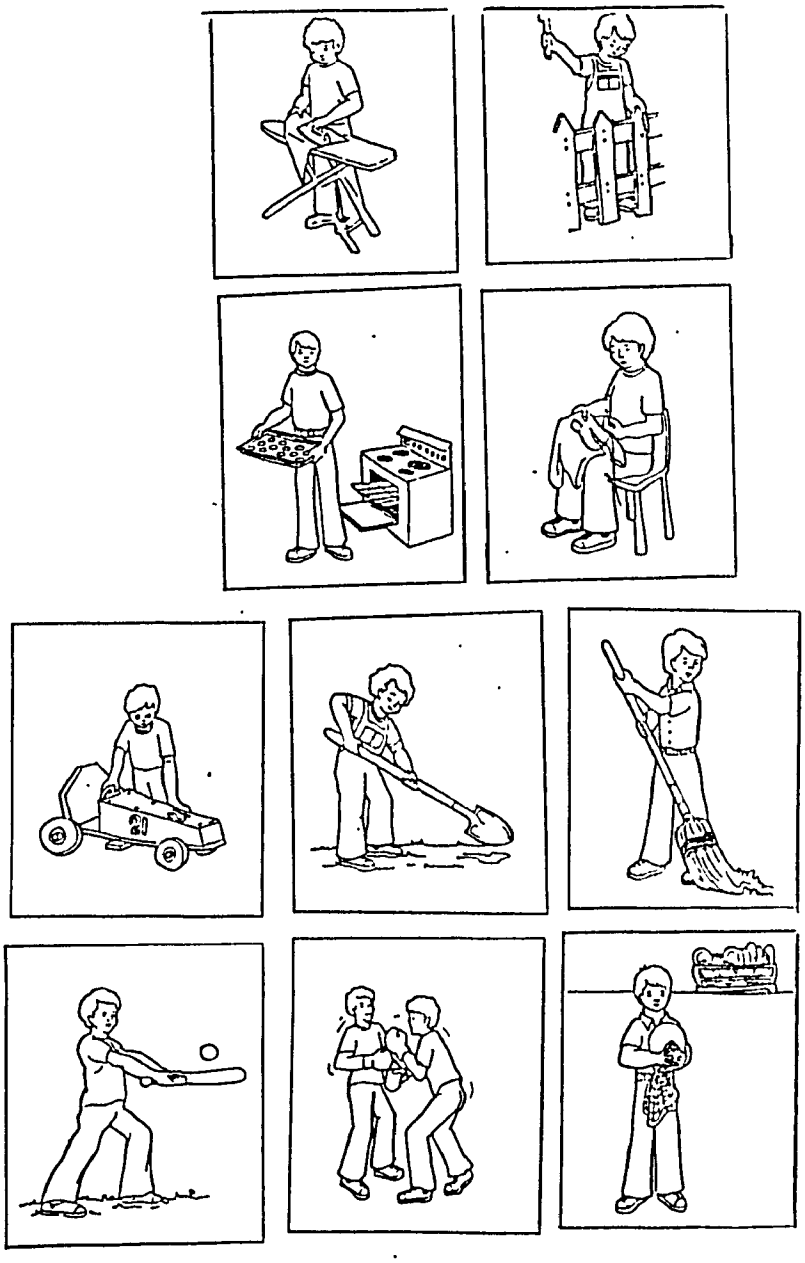
APPENDIX D (continued)



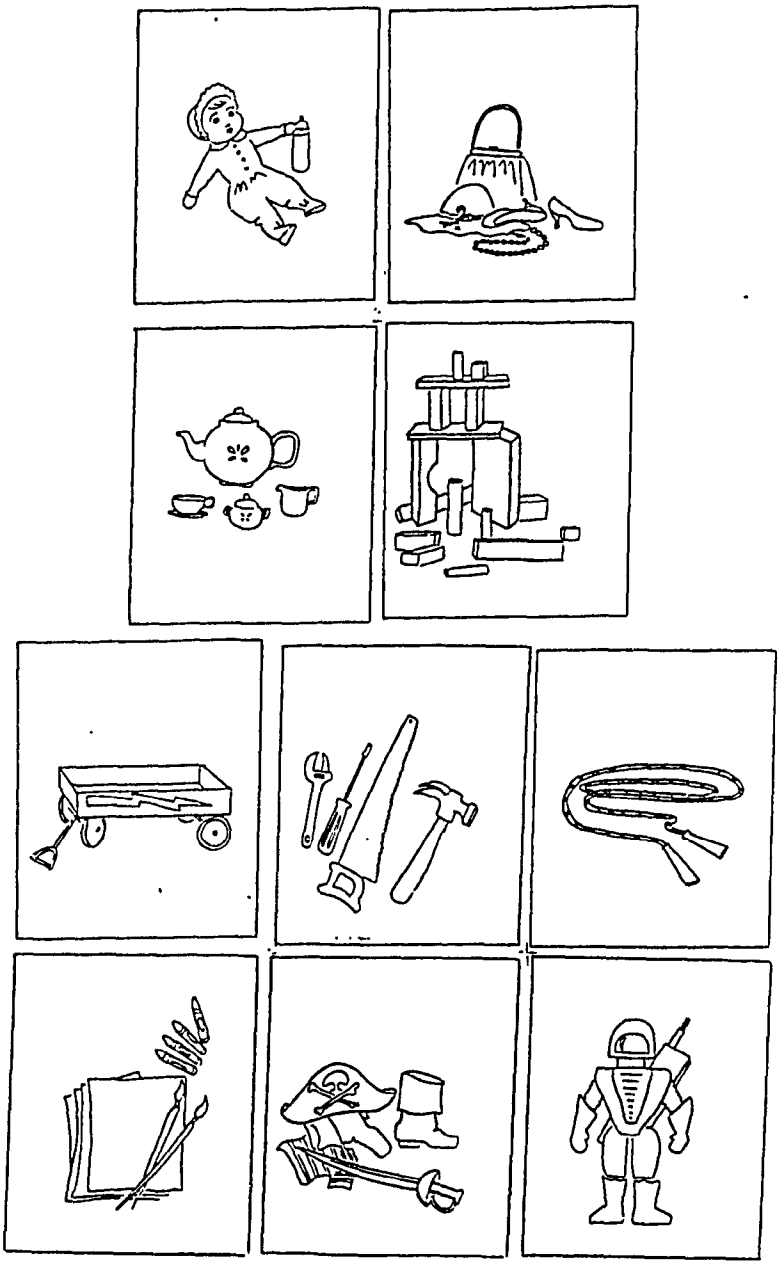
APPENDIX D (continued)



APPENDIX D (continued)



APPENDIX E
Pictures of Toy Items Included with SERLI



APPENDIX F

Source Table of ANOVA of Preference Scores by Age, Sex, Scores type (SRC SRP), and Activity Type (Child, Adult, Toy)

Source	SS	df	MS	F	p
WITHIN	29733.94	75	396.99		
AGE	2733.41	2	1366.71	3.44	.037 *
SEX	2235.90	1	2235.90	5.63	.020 *
AGE X SEX	645.79	2	322.89	.81	.447
WITHIN	23913.84	150	159.43		
ACTIVITY	1070.08	2	535.04	3.36	.038 *
AGE X ACTIV	790.92	4	197.73	1.24	.296
SEX X ACTIV	459.05	2	229.53	1.44	.240
AGE X SEX X ACTIVITY	1338.50	4	334.62	2.10	.084
WITHIN	2237.66	75	29.84		
SCORE	273.78	1	273.78	9.18	.003 *
AGE X SCORE	333.60	2	166.80	5.59	.005 *
SEX X SCORE	21.70	1	21.70	.73	.396
AGE X SEX X SCORE	47.90	2	23.95	.80	.452
WITHIN	6454.37	150	43.03		
ACTIV X SCORE	38.58	2	19.29	.45	.640
AGE X ACTIV X SCORE	55.82	4	13.85	.32	.861
SEX X ACTIV X SCORE	142.64	2	71.32	1.66	.194
AGE X SEX X ACTIV X SCORE	226.97	4	56.74	1.32	.266