
Genital Knowledge and Gender Constancy in Preschool Children

Author(s): Sandra Lipsitz Bem

Source: *Child Development*, Vol. 60, No. 3 (Jun., 1989), pp. 649-662

Published by: [Wiley](#) on behalf of the [Society for Research in Child Development](#)

Stable URL: <http://www.jstor.org/stable/1130730>

Accessed: 24-02-2016 16:36 UTC

REFERENCES

Linked references are available on JSTOR for this article:

http://www.jstor.org/stable/1130730?seq=1&cid=pdf-reference#references_tab_contents

You may need to log in to JSTOR to access the linked references.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Society for Research in Child Development and *Wiley* are collaborating with JSTOR to digitize, preserve and extend access to *Child Development*.

<http://www.jstor.org>

Genital Knowledge and Gender Constancy in Preschool Children

Sandra Lipsitz Bem

Cornell University

BEM, SANDRA LIPSITZ. *Genital Knowledge and Gender Constancy in Preschool Children*. CHILD DEVELOPMENT, 1989, 60, 649–662. This article challenges the widely cited view that the cognitive-developmental level of preschool children prevents them from conserving sex across perceptual transformations—from attaining “gender constancy.” Shortcomings in the procedures previously used to assess gender constancy are reviewed, and an empirical study is reported which uses both a new measure of gender constancy and a new test of a child’s genital knowledge. It was found that 40% of 3-, 4-, and early 5-year-old children could conserve sex across perceptual transformations, but only if they had the domain-specific knowledge that the genitalia constitute the defining attributes of male and female. Girls had significantly more genital knowledge than boys and displayed more gender constancy. A possible link between a child’s genital knowledge and his or her gender traditionalism is discussed.

Sixty years ago, Freud (1925/1959) asserted that the young child’s discovery of the genital difference between the sexes had profound motivational significance. Forty years later, Kohlberg (1966) disputed this emphasis on genital knowledge, arguing on the basis of Piaget’s stage theory of cognitive development that, until age 6 or 7, children are not fully capable of understanding that the genitalia constitute the defining attributes of male and female.

According to Piaget, children between approximately 18 months and 7 years are in a “preoperational” stage of cognitive development, a stage characterized by several intrinsic limitations of thought. For example, preoperational children are “perception bound,” that is, focused on the perceptual and spatial properties of an object and predisposed to treat these properties as defining even when they are not. In addition, they cannot yet “reverse” in thought a transformation that has been performed on some object in the real world and, hence, they are unable to “con-

serve” an object’s basic identity—to apprehend its underlying invariance—across perceptual transformations. (For more detailed summaries of Piagetian theory, see Flavell, 1963; Piaget, 1970.)

Two related aspects of children’s thinking about sex and gender follow theoretically from these characteristics of preoperational thought. Because preoperational children are perception bound, they define male and female in terms of visually salient perceptual cues like hairstyle and clothing rather than in terms of genitalia. Because preoperational children lack reversible thought, they are unable to apprehend that, even when a person’s outward appearance changes, his or her sex remains constant. This inability to conserve sex across perceptual transformations is described in the developmental literature as a lack of “gender constancy.”¹

During the past 20 years, a large number of empirical studies have confirmed that preschool children do indeed have difficulty on

The writing of this article and the empirical research it reports were supported by a grant from the Changing Gender Roles Program of the Rockefeller Foundation. I would like to thank both the parents and the staff at the Cornell Nursery School, the Ithaca Child Care Center, and the Ithaca Area Child Care Center for allowing us to interview their preschoolers. I would also like to thank Pamela Melton, Amy Gross, Martha Montes, Rachel Sifrey, Lindsey Bergman, Craig Wishman, Patricia Harney, and Patricia Peters for serving as research assistants during various phases of this project. Requests for reprints should be sent to Sandra L. Bem, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853.

¹ In general, this article follows the convention adopted by anthropologists of using the word “sex” to denote biological differences between males and females and the word “gender” to denote the cultural elaboration of those differences. Although the term “gender constancy” is not compatible with this usage—it is sex that is conserved, not gender—it is used here because of its familiarity to developmental psychologists.

[*Child Development*, 1989, 60, 649–662. © 1989 by the Society for Research in Child Development, Inc. All rights reserved. 0009-3920/89/6003-0007\$01.00]

gender-constancy tasks (e.g., DeVries, 1969; Emmerich, Goldman, Kirsh, & Sharabany, 1977; Gelman, Collman, & Maccoby, 1986; Kohlberg, 1966; Marcus & Overton, 1978; McConaghy, 1980; Slaby & Frey, 1975; Wehren & DeLisi, 1983). As we shall see, however, both theoretical and methodological considerations suggest that this difficulty may be more an artifact of faulty assessment procedures than a fact of early childhood.

Recently, a number of psychologists and philosophers working outside the domain of sex and gender have challenged the existence of discrete Piagetian-like stages. From their perspective, the reason young children fail to perform competently on a whole array of Piagetian tasks—including conservation tasks—is not that the children are preoperational thinkers, but that (a) the tasks are not well enough designed to tap just the particular cognitive capacity under test, and (b) the children do not yet have enough domain-specific knowledge about the subject matter being tested. (For empirical evidence consistent with this reasoning, see Carey, 1985; Chi, 1978; Gelman & Baillargeon, 1983; Gelman & Gallistel, 1978; Keil, in press.)

The purpose of this article is to test both of these propositions in the domain of sex and gender: Can preschool children conserve sex across perceptual transformations if they are assessed with a more carefully conceived test of gender constancy? Does their gender constancy depend upon their domain-specific knowledge that the genitalia constitute the defining attributes of male and female?

Previous Measures of Gender Constancy

Two measures have traditionally been used in the empirical research on gender constancy, the first designed by Slaby and Frey (1975), the second by Emmerich et al. (1977).

The Slaby-Frey Measure

The Slaby-Frey (1975) measure assesses the child's understanding of his or her own gender constancy by asking the following three questions: "If you wore [opposite-sex] clothes, would you be a girl or a boy?" "If you played [opposite-sex] games, would you be a girl or a boy?" "Could you be a [opposite sex] if you wanted to be?"

The first problem with the Slaby-Frey questions is that they can be interpreted in more than one way. Slaby and Frey clearly mean to ask, "If you temporarily put on cross-gender clothes or temporarily played with cross-gender toys, would your sex be trans-

formed in some very deep sense, or would it remain constant?" But the questions could also be interpreted by the child to mean, "Would you look like a girl or a boy?" or "Would you be pretending to be a girl or a boy?" or even "If you were the kind of child who regularly did [opposite-sex] things, probabilistically, is it likely that you would really be a girl or a boy?" Unfortunately, the correct answer to all of these alternative interpretations is the same as the incorrect answer to the Slaby-Frey interpretation. That is, I'd "be" the opposite sex.

The second problem with the Slaby-Frey questions is that they do not actually test the child's ability to conserve sex across a perceptual transformation, and hence they are inadequate as a test of Piagetian conservation. This, according to various critics (e.g., Marcus & Overton, 1978), is why as many as 40% of preschoolers are typically able to pass the Slaby-Frey measure of gender constancy but not other measures. This is also why other measures of gender constancy were ultimately developed, the most popular of which is the one designed by Emmerich et al. (1977).

The Emmerich Measure

The Emmerich (1977) measure presents the subject with a schematic drawing of a child with the clothing and hairstyle of either a boy or a girl. The drawing is in a notebook with a perfectly matched opposite-sex drawing positioned in exactly the same way on a page underneath the first. The top drawing is split at the neck so that either the head or the body or both can be removed to reveal an opposite-sex head or body or both on the page below. By removing either some portion or all of the top drawing, the interviewer can bring about either a partial or a complete gender transformation.

The Emmerich interview with these materials proceeds as follows. The interviewer presents the top drawing, provides a sex-appropriate name for the drawing, and announces the drawing's sex ("This is Janie; Janie is a girl"). To assess gender constancy, the interviewer removes part or all of the top drawing and then asks the child questions of the following form: "If Janie has her hair cut short like this and wears boys' clothes like this, what would she be? Would she be a girl, or would she be a boy?"

In their initial study of gender constancy in a group of economically disadvantaged children, Emmerich et al. (1977) reported that only 7% of 4-year-olds, 12% of 5-year-olds, 16% of 6-year-olds, and 25% of 7-year-olds

were able to conserve sex across perceptual transformations. In more middle-class samples, the percentage of conservers has typically been a bit higher, but not much higher (e.g., 23% vs. 16% for two samples of children with a mean age of 68 months, as calculated by Martin & Halverson, 1983). The widely cited conclusion that preschool children are incapable of gender constancy derives primarily from studies employing the Emmerich task.

The major problem with the Emmerich measure is the artificiality of its stimuli. When Emmerich and his colleagues ask a child whether "Janie" would be a girl or a boy if she had her hair cut short or wore boys' clothes, they clearly mean for the child to treat the original stimulus ("Janie") as if it were a real human being with real sexual invariance. But the child could just as reasonably treat the original stimulus as nothing more than a schematic drawing that the interviewer has defined as male or female purely on the basis of its hairstyle and clothing, in which case the correct response is that, if the drawing's hairstyle and clothing have changed, then the drawing's sex has changed as well. This is the "wrong" answer, of course, and hence the child is classified as a nonconserver.

Two recent studies have managed to circumvent some of these methodological shortcomings by using both modified instructions and human stimuli. To examine the child's understanding of gender constancy in the self, the first of these studies asked questions of the following form: "If you wore [opposite-sex] clothes, what would you *really* be, a boy or a girl?" The subjects were only 4, 5, and 6 years old, but a full 95% responded correctly (Martin & Halverson, 1983). To investigate the child's understanding of gender constancy in others, the second study took photographs of some of the subject's classmates dressed up in cross-sex costumes and then asked whether the child in each photograph was a boy or a girl. Although the subjects were only 3, 4, and 5 years old, virtually 100% responded correctly (Miller, 1984). These procedures are far from perfect. But like the more sensitive procedures designed by anti-stage researchers outside the domain of gender, they do suggest that existing measures of gender constancy may all seriously underestimate the preschool child's ability to conserve.

Where Is Biological Knowledge?

The concept of gender constancy would seem necessarily to require that a child understand that each individual has an underly-

ing biological "essence" of maleness or femaleness that remains invariant across surface transformations. Yet there are no studies using the Slaby-Frey or the Emmerich procedure that probe either the child's knowledge of or theories about the biology of maleness and femaleness. Some studies do ask children to explain their answers, of course, but none of the coding categories specifically taps the child's comprehension of what constitutes biological sex. An example of a good sex-conserving explanation is thus, "Because he was born a boy and will always be a boy." But what about, "Because he still has a penis under his clothes, not a vagina"? Does any child ever say such a thing? Does any child have this kind of anatomical knowledge? If a child did have such knowledge, would he or she have the nerve to verbalize it in a context where the adult in charge has done nothing to indicate that such talk is acceptable?

The only study in the literature that has ever dealt explicitly with the link between biological knowledge and gender constancy was done in Sweden (McConaghy, 1979, 1980). Working with a "gender jigsaw puzzle" that displayed schematic line drawings of two nude children who had no facial features whatsoever and who differed only with respect to hair length, clothing, and genitalia, McConaghy found that only 0% of 4-year-olds, 5% of 5- and 6-year-olds, 23% of 7-year-olds, and 34% of 8-year-olds appeared to understand that the male and female figures in her puzzle would remain male and female even if their (transparent) clothing or their hair underwent a gender transformation. Conceptually, this was clearly a step in the right direction, but unfortunately, McConaghy's procedures had many of the same methodological shortcomings noted earlier. Like Slaby and Frey's constancy questions, McConaghy's constancy questions were susceptible to more than one interpretation; and, like Emmerich's schematic drawings, McConaghy's schematic drawings (including the genitalia) were not at all realistic.

McConaghy's low percentages are striking because Swedish children are known to have much more biological knowledge about sex than American children (Goldman & Goldman, 1982). But even in those few studies where large numbers of preschool children do show gender constancy, it is simply not warranted to assume that the children have any real understanding of the biological basis of the sexual invariance that they appear to perceive. For example, in the Martin and Halverson (1983) and Miller (1984) studies,

the children may simply have been applying a well-rehearsed label to the self or to a familiar face without having the kind of biological knowledge that would enable them to conserve sex across a real gender transformation. In studies that have asked children to explain their answers (e.g., Emmerich et al., 1977; Wehren & DeLisi, 1983), the phenomenon of children simulating gender constancy is known as "pseudoconstancy" and seems primarily to involve children who ignore or deny the hypothetical premise upon which the interviewer's question is based. Shown a picture of Janie wearing boys' clothes, the pseudoconstant child thus asserts that Janie is a girl because "girls don't wear boys' clothes."

The current study was designed to investigate the child's knowledge about genital sex differences and the relation of that knowledge to gender constancy. Genital knowledge was emphasized rather than, say, reproductive or chromosomal knowledge both because the genitals are anatomically visible and also because preschoolers seemed more likely to be familiar with them. In principle, however, knowledge of any biological invariance should foster an understanding of gender constancy.

Genital Knowledge and Gender Constancy: An Empirical Study

As noted earlier, both a new gender-constancy measure and a new genital knowledge test were designed for the current study. The stimuli for both were 8 × 10, full-length, color photographs of toddlers between the ages of 1 and 2. The toddlers were nude from at least the waist down and had wispy, baby-like hair that did not demarcate sex. Color photographs were used rather than schematic drawings so that a real person—with real sexual invariance—would be represented in the stimulus materials.²

As we shall see, conservation was assessed by first showing the child a photograph of a particular toddler in the nude and then determining whether the child could correctly identify that same toddler's sex when he or she was seen a moment later fully clothed and coiffed in a gender-inconsistent way and, then, in a gender-consistent way. Both a particular male toddler and a particular female toddler were seen in these three states. A child was defined as showing con-

servation if he or she correctly identified the sex of both toddlers in both their gender-inconsistent and their gender-consistent states. Note that children who show conservation are ignoring the cultural cues of hairstyle and clothing and are focusing instead on the now-concealed biological cue of genitalia.

In order to avoid the possibility that repeated exposure to nude toddlers might itself facilitate conservation by priming children to attend to the genital difference between the sexes, the gender-constancy measure was given before the genital knowledge test, and the subtests of the genital knowledge test were themselves sequenced from hardest to easiest. Also, no feedback was ever given to subjects about the correctness of their answers. These methodological features all work against our discovery of early gender constancy.

Method

Subjects

Fifty-eight children between 36 and 65 months of age participated as subjects. Of these, 31 were girls and 27 were boys. Each child was tested individually in a private room of their nursery school or day-care center by one of six female interviewers. Some children completed the full test schedule in one session; others required as many as three sessions. For two nursery school groups and one day-care group, the testing occurred during 1986–1987; for the second day-care group, the testing occurred during 1985–1986.

Because of the sensitive nature of the stimulus materials, special care was taken to fully inform parents before they consented to have their children participate in the research. Even so, 83% of the parents gave their consent.

The Gender-Constancy Measure

The stimuli for the gender-constancy measure were six nude or seminude photographs, three taken of a particular male toddler, three taken of a particular female toddler. The three photographs of each toddler were all contained within a single manila folder; the folder had two inside pockets that covered the photographs up to the waist. The first photograph in each set was fully nude; the second photograph was gender-inconsis-

² Each photograph was taken in the toddler's own home with at least one parent present. Parental consent was obtained in writing to use the photographs in empirical research on children's thinking about maleness and femaleness. In the case of "Gaw" and "Khwan," parental consent was also obtained to publish the photographs.

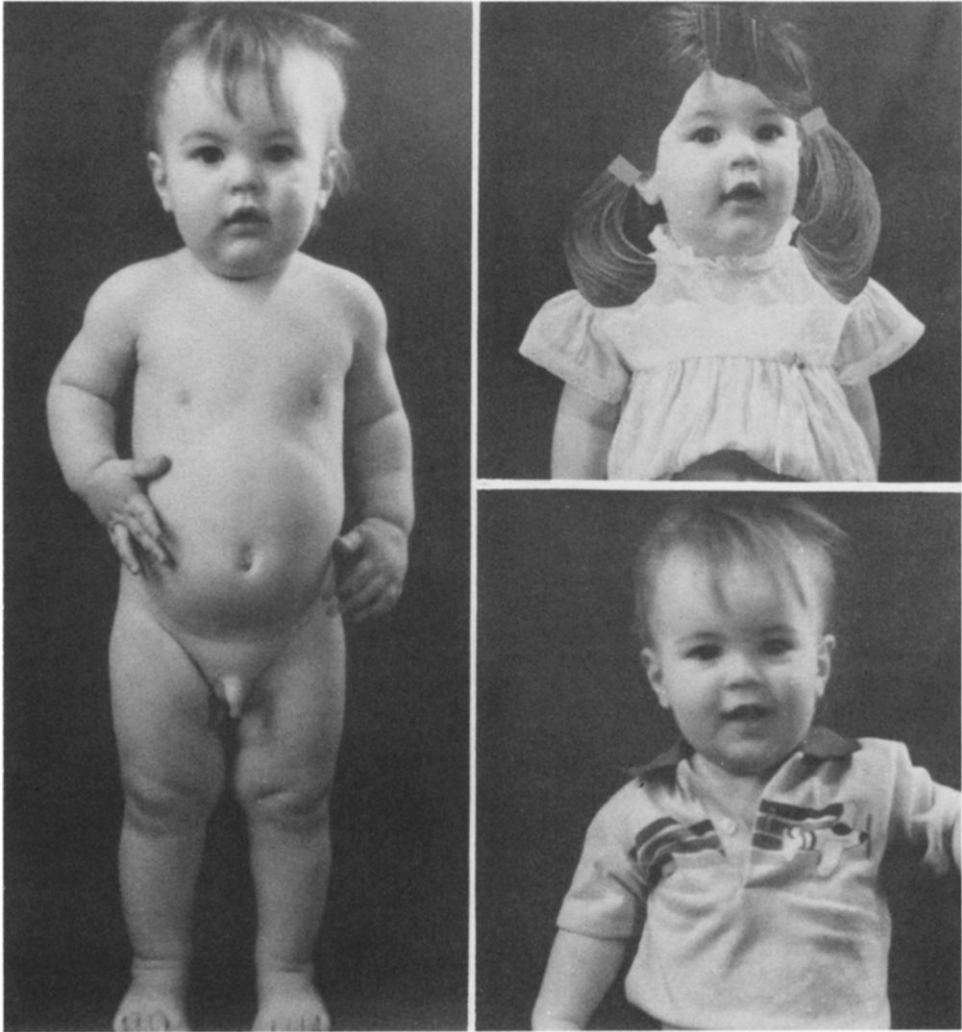


FIG. 1.—The male photographs used in the sex conservation test

tent (the boy had long ponytails with pink barrettes and a frilly pink blouse; the girl had a blue and gray striped polo shirt and a football); the third photograph was gender-consistent (the boy had the polo shirt and no ponytails; the girl had the frilly pink blouse, a purse, and a lipstick). The gender-consistent and gender-inconsistent photographs were always displayed from inside the pockets of their respective folders so that the genitals were never visible. The nude photographs were always displayed by removing them from the folders so that the genitals were clearly visible. The three photographs from each set are displayed in Figures 1 and 2. Note that although even girl toddlers do not ordinarily have hair as full or as long as the ponytails on the boy in Figure 1, the children

did not find the hairstyle anomalous because, as we shall see shortly, a plausible rationale was provided for the boy suddenly looking so much like a girl, and the hair itself was explicitly described as a wig.

After the child had been told explicitly that the three photographs in the first folder were "all pictures of the same baby," a baby named "Gaw," the interviewer asked the child a specified set of questions first about the nude photograph of Gaw, then about the gender-inconsistent photograph, and finally about the gender-consistent photograph. (The unfamiliar Thai names of "Gaw" and "Khwan" were used so that they would not provide cues to the toddlers' sex. Gaw is the boy, Khwan, the girl.)

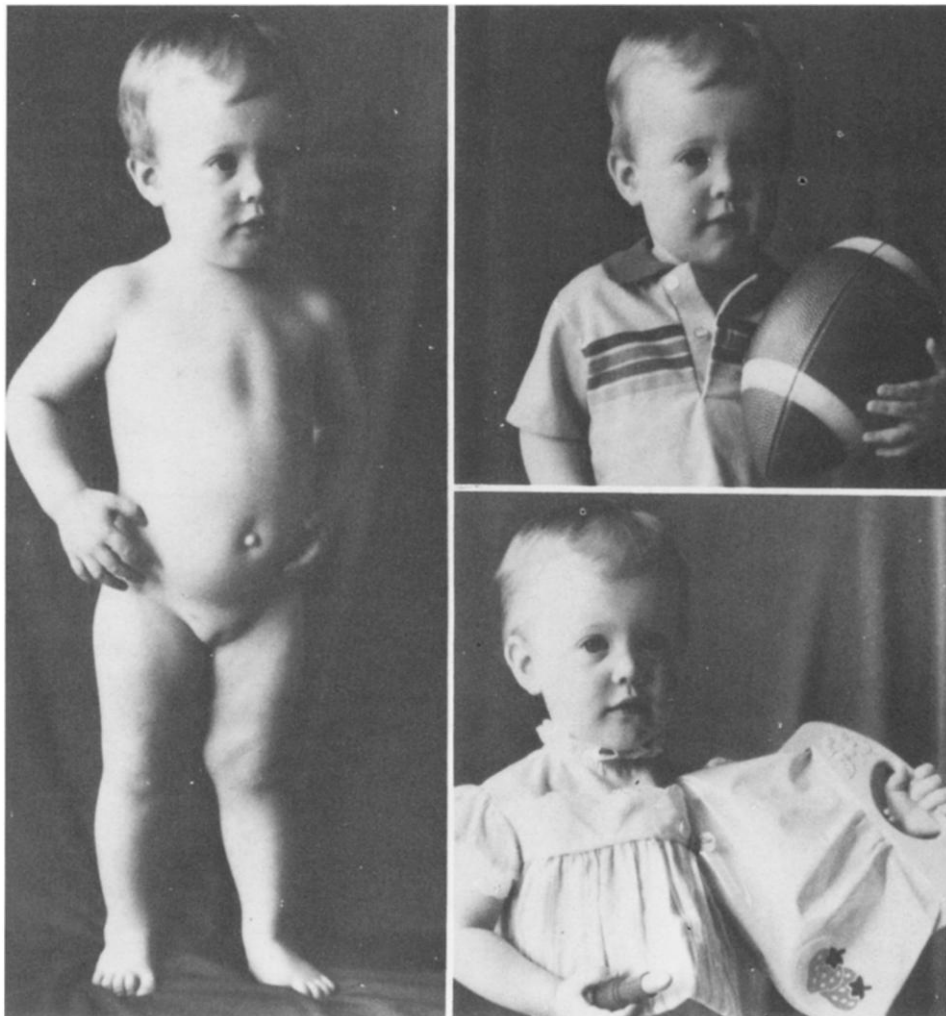


FIG. 2.—The female photographs used in the sex conservation test

The nude photograph.—The interviewer removed the photograph from the folder, told the child to look at it very carefully, and then asked the child whether Gaw was a boy or a girl. After the child responded, the interviewer probed for the basis of the child's judgment with the following series of increasingly directive questions, the later questions being asked only if the earlier questions yielded no articulated genital knowledge: (a) "How do you know Gaw is a [boy][girl]?" "What makes Gaw a [boy][girl]?" (b) "Is there anything about Gaw's body that makes Gaw a [boy][girl]?" (c) "Can you point to anything on Gaw's body that makes Gaw a [boy][girl]?" (d) "Does that part of Gaw's body have a name?"

The gender-inconsistent photograph.—The interviewer first provided the child with

a rationale for Gaw's gender transformation by saying that she had a picture of Gaw taken "one day when Gaw was playing silly dress-up games and got all dressed up in a girl's wig and a girl's blouse." With the nude photograph of Gaw still visible, the interviewer then opened the folder to reveal the gender-inconsistent picture of Gaw displayed from inside its pocket. After the child had been asked whether he or she could see that it was "still the same baby," that it was "still Gaw," the nude photograph was removed from sight, the child was asked to look very carefully at the gender-inconsistent photograph, and the following three questions were asked: (a) "What does Gaw look like—a boy or a girl?" (b) "What is Gaw really—a boy or a girl?" (c) "What makes Gaw really a [boy][girl]?" If the child made no spontaneous mention of genitalia while answering these three questions, a

fourth question was asked as well: (d) "Is there anything about Gaw's body that makes Gaw really a [boy][girl]?"

The gender-consistent photograph.—The interviewer told the child that she had "yet another picture of Gaw taken when Gaw was all finished playing dress-up, had taken off the girl's wig and the girl's blouse, and had put Gaw's own shirt back on." With both the nude and the gender-inconsistent photographs of Gaw visible, the interviewer then opened the folder still further to reveal the gender-consistent photograph of Gaw. After the child had again been asked whether he or she could see that it was "still the same baby," that it was "still Gaw," the nude and the gender-inconsistent photographs were removed from sight, and the child was asked the same set of questions described above for the gender-inconsistent photograph.

After all these questions had been asked about the three photographs of the boy Gaw, the same questions were asked about the three photographs of the girl Khwan. The Khwan script differed from the Gaw script in two ways. When introducing the gender-inconsistent photograph, the interviewer said that she had a picture of Khwan taken "one day when Khwan had put on a shirt and started playing with a football." When introducing the gender-consistent photograph, the interviewer said that she had "yet another picture of Khwan taken after Khwan put some different clothes on and started to play with a purse and some lipstick."

The two scripts differ deliberately because even 3-year-olds already know that males do not "cross-dress" or "cross-play" in our culture nearly so much as females do. At one level, the reason for this is obvious: Even by 2 and 3 years of age, there are already more social costs for male gender deviance than for female gender deviance (e.g., Fagot, 1977). Given this asymmetry in social reality, the psychological impact of the two cross-gender photographs would have been very different even if the two scripts had been identical. Rather than ignoring this difference, the two scripts acknowledge it by providing a more elaborate rationale for the gender-inconsistent Gaw photograph than for the gender-inconsistent Khwan photograph.

Social reality impinged on the Gaw and Khwan procedures in two other ways. First, only Gaw's hairstyle was ever transformed because, like many girl toddlers, Khwan had short "boyish" hair to begin with. Second, although Gaw's gender-inconsistent photograph was explicitly characterized in gender

terms—"Gaw . . . got all dressed up in a girl's wig and a girl's blouse"—Khwan was never characterized as doing anything explicitly male. Again, this is because girls and women cross-dress and cross-play so much more than boys and men in everyday life that many more things are stereotyped as exclusively female.

These several procedural differences had no detectable impact on the children's reactions to the two sets of pictures. There was no commentary indicating that they found one set any more noteworthy than the other. There was also no more sex conservation on one set than on the other.

It should also be noted that those subjects tested during 1985–1986 actually took an earlier version of the gender-constancy measure that (a) transformed the two toddlers' clothing and toys but not the male toddler's hairstyle, and (b) asked what sex the toddlers "really were" but not what sex the toddlers "looked like." Because the data from the preliminary and final versions were virtually identical, they were combined into a single data set.

Only one question was considered in the scoring of the conservation test: "What is Gaw [or Khwan] really—a boy or a girl?" The criterion for passing the conservation test was correctly answering this question all four times it was asked, that is, correctly identifying the real sex of *both* the gender-inconsistent Gaw and Khwan *and* the gender-consistent Gaw and Khwan. Note that although we first asked children what sex Gaw and Khwan looked like and only then asked what sex Gaw and Khwan really were, we were not thereby enabling children to pass the conservation test merely by switching the toddler's sex in response to what might be seen as the demand characteristics of the second question. This is because the toddlers really are a different sex from the one they look like only in the gender-inconsistent photographs. In the gender-consistent photographs, the toddlers really are the same sex as the one they look like.

The Genital Knowledge Test

The genital knowledge test asked children to identify the sex of two male and two female toddlers other than Gaw and Khwan. The photographs of these four new toddlers were arranged into four subtests, all of which always showed the toddlers with their genitals clearly visible. The children were not informed that the same four toddlers would be appearing more than once.

On the first (and most difficult) subtest, 16 photographs were presented one at a time,

and the child was asked about each whether it was a boy or a girl. For the first four photographs only, the child was also asked the basis for his or her judgment: "How do you know it's a [boy][girl]?" "What makes it a [boy][girl]?" Each of the four toddlers thus appeared four times overall, once wearing a "masculine" shirt, once wearing a "feminine" shirt, once holding a "masculine" toy, and once holding a "feminine" toy. Each shirt and toy appeared twice, once with a male toddler and once with a female toddler. Although every toddler thereby appeared with both gender-consistent and gender-inconsistent cultural cues, only the gender-inconsistent photographs were scored. A child was said to pass the first subtest if he or she correctly identified the toddler's sex on seven of the eight gender-inconsistent photographs.

On the second subtest, these same eight gender-inconsistent photographs were presented again. This time, however, a male and a female were presented side by side, and the child was asked which was the boy and which the girl. A child was said to pass the second subtest if he or she correctly identified both the boy and the girl on all four pairs.

Note that these first two subtests assess the child's knowledge that genital cues take priority over cultural cues, and they are thus conceptually quite similar to the gender constancy measure. They are much easier, however, because both the cultural and the genital cues are clearly visible at the critical moment when the child is asked the toddler's sex. In contrast, only the misleading cultural cues are visible at that critical moment on the gender-constancy test.

The next two subtests simply assess the child's knowledge of the genital difference between the sexes. On the third subtest, fully nude photographs of all four toddlers were presented one at a time, and the child was asked whether the toddler was a boy or a girl. The child's earlier identification of the sex of both the nude Gaw and the nude Khwan was also counted in the scoring, however, thereby constituting a six-item rather than a four-item test. A child was said to pass the third subtest if he or she correctly identified the toddler's sex in five out of the six photographs.

On the fourth and final subtest, the nude photographs of the four new toddlers were presented again, but this time in male-female pairs. A child was said to pass the fourth subtest if he or she correctly identified both the boy and the girl on both nude pairs.

It should be noted that all the shirts and toys used in both the gender-constancy measure and the genital knowledge test were first judged to be either masculine or feminine by the author and two undergraduate research assistants. They were subsequently judged again by the subjects themselves. More specifically, at the conclusion of each child's final session, he or she was asked to look at a series of 32 color photographs of toys and clothes and to indicate about each item "who plays with it" or "wears it more, boys or girls?" Included in this series were the 12 shirts and toys that had been selected for the gender-constancy and genital knowledge tests. Of these 12, 11 were judged as either masculine or feminine by at least 80% of the preschoolers, and the twelfth was so judged by 63% of the preschoolers.

Other Measures

In addition to the gender-constancy and genital knowledge tests, two other tests were administered as well: (1) a cultural gender test, which was administered merely to reconfirm that even very young children can distinguish males from females on the basis of cultural cues; and (2) an own-sex interview, which assessed the children's ability to conserve their own sex. For all children, the cultural gender test was administered first, the gender-constancy measure second, and the genital knowledge test last. Only those children tested during 1986–1987 took the own-sex interview; it was administered between the gender-constancy measure and the genital knowledge test. For this subset of children only, there were thus two measures of gender constancy, one involving others (i.e., toddlers in photographs) and one involving the self.

The cultural gender test.—Ten 3 × 5, head-and-shoulder, color photographs of school-age children and adolescents clothed and coiffed in a traditionally sex-typed way were presented one at a time, and the subject was asked whether the model in the photograph was a boy or a girl. Half the models were boys and half were girls. All the photographs had originally been taken from various mail-order shopping catalogs by Leinbach and Fagot (1986) for use in their research. Overall, the preschool children in the current study correctly identified 98.4% of these catalog pictures, thereby reconfirming that they can indeed distinguish males from females on the basis of cultural cues alone. Because of the virtually perfect performance by all children, this measure will not be discussed further.

The own-sex interview.—The own-sex interview asked children the following three questions adapted from the Gaw-Khwan constancy measure:

1. Remember how Gaw was playing silly dress-up games with a girl's wig and a girl's blouse? Well, if *you* were playing silly dress-up games one day and you put on [TO BOYS: a girl's wig with real real long hair and girls' clothes; TO GIRLS: a boy's wig with real real short hair and boys' clothes], what would you look like? Would you look like a boy or would you look like a girl?

2. Now I want to ask you something just a little bit different. If you put on [TO BOYS: a girl's wig with real real long hair and girls' clothes; TO GIRLS: a boy's wig with real real short hair and boys' clothes], what would you *really* be? Would you really be a boy or would you really be a girl? Why would you really be a [boy][girl]?

3. Could you really be a [opposite sex of child] if you wanted to be? If YES: How? If NO: Why not?

Results

Genital Knowledge and Gender Constancy

Overall, 40% of the 3-, 4-, and early 5-year-old children in this study were able to conserve sex across perceptual transformations. Even more importantly, only those children who had domain-specific knowledge about both the genital difference between the sexes and the priority of genital cues over cultural cues were able to do so. Two separate analyses support this conclusion.

First, a scalogram analysis on the pass-fail pattern of children's subtest performances revealed that the four subtests of the genital knowledge test and the Gaw-Khwan conservation test did form an ordered sequence. As can be seen in Table 1, 51 of the 58 children—or 88%—showed one of the within-scale patterns, and only seven children violated that scale (reproducibility index = +.97; consistency index = +.83; Green, 1956). Success on more difficult subtests of the genital knowledge test thus requires the knowledge needed for success on easier subtests. And conservation, in turn, requires both the knowledge of male and female genitalia and the knowledge that, whenever genital and cultural cues are in conflict, genital cues have priority in determining sex.

Second, of those 31 children who passed all four of the genital knowledge subtests, thereby indicating that they had the domain-specific knowledge required for conservation, fully 23 children—or 74%—passed the conservation test as well; in contrast, of those 27 children who did not pass all four of those

subtests, only 11% passed the conservation test, $\chi^2(1) = 23.2$, $p < .001$. Because these three children were all very young and had virtually no genital knowledge at all, it seems likely that they passed the conservation test by chance. Of those 23 children who passed both the conservation test and the genital knowledge subtests, eight were under 4 years of age, with the youngest being only 37 months.

In sum, preschool children as young as 37 months can conserve sex on our new sex conservation test. Moreover, they can do so only if they have the domain-specific knowledge that the genitalia are the defining attributes of sex and that the characteristic cultural markers of sex are not.

Genital knowledge is not the entire story, however, for there was still an age effect in our data: In particular, among children who had genital knowledge, more 4- and early 5-year-olds passed the conservation task than 3-year-olds: 88% and 57%, respectively, $\chi^2(1) = 3.88$, $p < .05$. Those 3-year-olds who did pass, however, were not any older than those who did not pass.

Sex Differences in Biological Knowledge and Genital Language

For purposes of further analysis, each child was assigned a single genital knowledge score equal to the total number of tests that he or she passed. As can be seen in the upper-left portion of Table 1, this genital knowledge score could range from 0 to 5. These scores were analyzed by means of a sex \times age ANOVA using the method of unweighted means to adjust for the unequal number of subjects per cell (Ferguson, 1981). Table 2 presents the mean genital knowledge scores for the two sexes and the two age groups in this sample.

As can be seen in the table, there were significant main effects for both age, $F(1,54) = 9.12$, $p < .01$, and sex, $F(1,54) = 12.01$, $p < .01$, with both older children and females having significantly more genital knowledge than younger children and males, respectively. But there was also a highly significant interaction between age and sex, $F(1,54) = 9.84$, $p < .01$, indicating that it was actually only among males that genital knowledge increased with age. Females, in contrast, had as much genital knowledge at age 3 as either males or females had by age 5. This sex \times age interaction can also be seen in the fact that the seven youngest conservers were all girls.

It is not clear why the 3-year-old girls in this sample had so much more biological

TABLE 1
A GUTTMAN SCALE OF GENITAL KNOWLEDGE

PATTERN	SUBTEST					Boys (<i>n</i> = 27)		Girls (<i>n</i> = 31)	
	Nude Pairs	Nude Singles	Inconsistent Pairs	Inconsistent Singles	Constancy	<i>N</i>	Mean Age	<i>N</i>	Mean Age
Within-scale:									
0.....	-	-	-	-	-	11	3-7	3	3-10
1.....	+	-	-	-	-	2	4-4	3	3-7
2.....	+	+	-	-	-	2	4-1	1	4-1
3.....	+	+	+	-	-	1	4-0	1	3-4
4.....	+	+	+	+	-	1	5-4	6	3-6
5.....	+	+	+	+	+	6	4-9	14	4-2
Total within-scale						23		28	
Violation:									
A ^a	-	-	-	-	+	2	3-4	0	...
B ^b	-	+	-	-	+	0	...	1	3-6
C ^c	+	+	-	+	-	0	...	1	3-9
D ^d	+	+	+	-	+	1	5-1	0	...
E ^d	+	+	-	+	+	1	4-0	1	3-4
Total violation						4		3	

^a Reassigned to - - - - -.

^b Reassigned to + + - - -.

^c Reassigned to + + + + -.

^d Reassigned to + + + + +.

TABLE 2

MEAN GENITAL KNOWLEDGE SCORES
BY AGE AND SEX

SEX	AGE	
	3 (<i>n</i> = 32)	4 and 5 (<i>n</i> = 26)
Boys (<i>n</i> = 27).....	.57	3.46
Girls (<i>n</i> = 31)	3.67	3.62

NOTE.—Sex $F(1,54) = 12.02, p < .01$; age $F(1,54) = 9.12, p < .01$; interaction $F(1,54) = 9.84, p < .01$.

knowledge than the 3-year-old boys. One possibility, however, is that girls may have more opportunity than boys to learn about opposite-sex genitalia. Boys' genitals are more visible, after all. Moreover, males enjoy higher status in our society, which means that all their attributes—including their genitals—are more likely to be culturally named.

Consistent with this hypothesis is the pattern of genital language used by boys and girls, respectively. In particular, there was a highly significant sex difference in the expression of female genital terms, with 58% of the girls but only 15% of the boys saying "vagina" or some equivalent at least once during their interview, $\chi^2(1) = 11.47, p < .001$. In contrast, there was no such sex difference in the expression of male genital terms, with 67% of the boys and 68% of the girls at least once saying "penis" or some equivalent, $\chi^2(1) < 1$, N.S. (Penis and vagina were by far the most commonly used sex-specific genital terms in these data. Other male terms used were tinkle, z-z, peepee, tail, tushi, and peanuts; other female terms used were vulva, clitoris, ginny, asty, crack, and hynee.) Overall, then, the girls did know more than the boys about opposite-sex genitalia, which may partially account for their superior performance on the genital knowledge scale at age 3.

One other sex difference in genital language emerged as well. Among those few children who verbalized no genital terms at all, significantly more girls than boys were nevertheless able to identify toddlers as male or female on the basis of their genitalia ($p < .05$ by a Fisher exact probability test; Siegel, 1956). Table 3 displays the relevant data. As can be seen in the table, when the boys in our sample used no genital language, it was because they did not know much about genitalia. In contrast, when the girls used no genital language, there was a good chance that they knew more than they were willing to

TABLE 3

NUMBER OF BOYS AND GIRLS USING NO GENITAL
LANGUAGE WHO EITHER HAD NO GENITAL
KNOWLEDGE OR SOME GENITAL KNOWLEDGE

SEX	GENITAL KNOWLEDGE SCORE	
	No Knowledge (0)	Some Knowledge (1, 2, 3, 4, or 5)
Boys ...	6	1
Girls ...	1	4

NOTE.— $p < .05$ by a Fisher exact probability test.

verbalize. Although it cannot be seen in the table, two of the girls actually knew a lot more, having passed all five subtests, including the gender-constancy test. A third girl passed three subtests, and a fourth, two subtests.

Knowledge of One's Own Sexual Invariance

It will be recalled that 38 of the children in the current sample were also asked a series of questions about their own sexual invariance. Briefly, "If [like Gaw] you were playing silly dress-up games one day and you put on a girl's [boy's] wig . . . and girls' [boys'] clothes, what would you look like, a boy or a girl? What would you really be? . . . Why? . . . Could you really be a boy [girl] if you wanted to be? How?/Why not?"

On the basis of his or her responses to these several questions, two independent coders rated each child as either "clearly demonstrating" or "not clearly demonstrating" an understanding of sexual invariance in the self. Children were coded as clearly demonstrating sex conservation if (a) they consistently answered all three of the major questions correctly either with or without a clarifying explanation (i.e., I'd look like the opposite sex; I'd really be my own sex; I couldn't be the opposite sex even if I wanted to be); or (b) they provided a verbal explanation that readily resolved any apparent inconsistency in their answers to the three major questions (e.g., I'd look like my own sex because my long hair would be sticking out from under the boy's wig; I could be a boy if I wanted to be because girls and boys can play with the same toys and wear the same clothes, but I wouldn't have a penis). Note that, in contrast to the Gaw/Khwan measure of sex conservation, this measure does not specifically require any genital knowledge on the part of the child. The two coders agreed perfectly on 92% of the 38 subjects; their three disagreements were resolved by discussion.

A sex \times age analysis of variance on children's pass/fail scores for this measure of sex conservation in the self revealed no significant main effects of either sex or age and no significant interactions. Overall, 53% of the preschool children in the current sample appeared to understand that their own sex would stay the same even if their appearance underwent a gender transformation, which is somewhat more than the 40% who understood that both Gaw's and Khwan's sex would stay the same if their appearance underwent a gender transformation. The two measures were closely related, however, with 25 of the children in the current sample—or 66%—either passing or failing both, $\chi^2(1) = 3.98$, $p < .05$; Pearson $r = +.32$.

That still leaves 13 children who passed one measure while failing the other, of course. Especially illuminating are the eight of these who passed the self measure while failing the Gaw/Khwan measure. These eight subjects all seemed to be quite sure that they themselves were, always had been, and always would be either boys or girls. As they themselves put it, they are the sex they are and they cannot change because "that's how I was made"; "I was born a girl"; "Jesus made me a boy." At the same time, however, six of the eight could not manage to pass even one subtest of the genital knowledge test, which means that they could not tell which of two babies was a boy and which was a girl even when the two babies were totally nude and pictured side by side. Children like these suggest that—my emphasis on genital knowledge notwithstanding—there may well be more than one way for a child to understand sexual invariance.

Discussion

As noted in the introduction, the preschool child's difficulty with gender-constancy tasks was initially seen as following naturally from the child's preoperational stage of cognitive development. More recently, however, it was suggested that preschool children may do badly on gender-constancy tests not because they are intrinsically unable to overcome the lure of perceptually misleading information, but because all such tests require them to make "category" inferences (e.g., Is the child a boy or a girl given that it has both a penis and long hair?) rather than "property" inferences (e.g., Does the child have a penis or a vagina given that it is a boy with long hair?). Consistent with this suggestion is the finding that preschool children were better able to ignore conflicting perceptual informa-

tion about gender on a new property inference task than on either a new category inference task or on the Emmerich gender-constancy task (Gelman et al., 1986).

The results of the current study question both of these interpretations by demonstrating that fully 40% of 3-, 4-, and early 5-year-old children are capable of conserving sex across perceptual transformations on the new Gaw-Khwan measure—a category inference task—if they have the domain-specific knowledge that genital cues take priority over cultural cues in defining sex. Although this 40% figure is identical to that reported by Slaby and Frey (1975) for their purely verbal measure of gender constancy, it is much higher than the 0% to 12% typically reported for this age group on the two perceptual measures designed by Emmerich et al. (1977) and McConaghy (1979), respectively.

At first glance, it might seem that the Gaw-Khwan test is simply easier than previous tests. After all, it uses realistic color photographs rather than schematic drawings; it emphasizes that all the pictures are pictures of the same toddler; it provides a plausible rationale for the toddler's gender transformation; and it explicitly makes a distinction between what sex the toddler looks like and what sex the toddler really is.

But these procedural innovations do not in any way reduce the amount of domain-specific knowledge required to pass the test. Thus, in order to conserve on our measure, a subject still must be able: (a) to distinguish between males and females on the basis of genital cues alone, (b) to give priority to genital cues over cultural cues when both are visible and in conflict, and (c) to give priority to genital over cultural cues even when the genital cues are hidden from sight and only the misleading cultural cues are visible. If our test is easier than previous tests, it is so only in the sense that it is less artificial, less susceptible to alternative interpretations, and based on a finer-grained task analysis of what gender constancy actually is and what domain-specific knowledge actually underlies it.

The virtues of the Gaw-Khwan measure notwithstanding, however, 60% of the preschool children in the sample still failed to conserve. These 35 "misses" fell into two distinct groups.

Twenty-seven of them—or 77%—did not pass the four genital knowledge subtests, and hence they had too little domain-specific knowledge for conservation to be possible. It is not clear why these particular children had

so much less genital knowledge than other children. Fully two-thirds of these insufficient-knowledge children were boys, however, and, as noted earlier, preschool boys appear to know strikingly little about opposite-sex genitalia in particular.

The other eight "misses" all had sufficient knowledge for conservation; yet they still failed the gender constancy test. In part, the failure of this group is surely due to the fact that the conservation test is harder than any of the genital knowledge subtests. In part, however, the failure of this group is also surely due to the way we decided to sequence the procedures in this particular study.

Recall that in order to guard against the possibility that the genital knowledge subtests might themselves facilitate conservation by priming children to attend to genital cues, we gave children the gender constancy test before the genital knowledge subtests. As it happened, moreover, we probably also inadvertently primed the children's knowledge of cultural cues by asking them to identify the sex of stereotypically sex-typed models from mail-order catalogs immediately before giving them the gender constancy test. Although these sequencing decisions all worked against their being able to pass the conservation test, fully 74% of the knowledgeable children passed anyway, as noted earlier. Of the eight knowledgeable children who did not pass, six were 3-year-olds (all girls).

Genital Knowledge and Gender Traditionalism

Developmental psychology became interested in the phenomenon of gender constancy because of Kohlberg's (1966) implication that, by virtue of being preoperational thinkers, preschool children were perception-bound nonconservers predisposed to define male and female on the basis of visually salient cues like hairstyle and clothing. But Kohlberg's account also implied that preschool children were egocentric "moral realists" (Piaget, 1932) predisposed to treat all rules and regularities—especially those about gender—as absolute and universal. For Kohlberg, preoperational thought thus led directly to gender traditionalism: "the physical constancies underlying . . . gender . . . tend to be identified with divine or moral law, and the need to adapt to the physical realities of one's identity is viewed as a moral obligation" (Kohlberg, 1966, p. 122).

Elsewhere I have argued that gender traditionalism is not an inescapable characteris-

tic of early childhood (Bem, 1988). Here I should simply like to repeat a suggestion made there that an unambiguous genital definition of sex might itself begin to attenuate children's traditional gender traditionalism.

As I see it, children define sex on the basis of visually salient cultural cues like hairstyle and clothing simply because they have picked up an implicit—if somewhat erroneous—cultural metamessage about what sex is. As a culture, we dress males and females differently and give them different hairstyles precisely so that their sex will always be apparent even when their genitalia are hidden. In supermarkets, on playgrounds, and in every other social context, we also readily identify people as male or female for our children even while having no specific information about those people's genitalia. In doing these things, we adults are not only relying on visually salient cultural cues ourselves. We are also unwittingly communicating to our children that these cues are the defining attributes of male and female.

In principle, of course, we could communicate a quite different metamessage not only about what sex is, but also about when sex matters. Specifically, we could communicate that sex is a narrowly construed biological concept that does not need to matter very much outside the domain of reproduction, which is the antithesis of the traditional cultural metamessage that sex matters very much indeed in virtually all domains of human activity. The implication here, of course, is that children exposed to this alternative metamessage would be much less disposed to believe that every arbitrary gender rule is a moral absolute, every violation a disturbing redefinition of one's maleness or femaleness.

This is not to say that biological knowledge would necessarily liberate all children from gender traditionalism. By itself, in fact, biological knowledge could just as well legitimize the cultural emphasis on the male-female dichotomy, thereby making the child all that much more eager to behave in gender-stereotyped ways. But if biological knowledge came to the child as part of a larger lesson about how unimportant one's sex is outside the domain of reproduction, then the child could well conclude that he or she can disregard gender stereotypes. After all, the child could reason, it is my body, not my behavior, that makes me either a male or a female; hence I can behave in any way that I please.

Both the liberation that can come from having a narrow biological definition of sex

and the imprisonment that can come from not having such a definition are strikingly illustrated by an encounter my son Jeremy had the day he naively decided to wear barrettes to nursery school. Several times that day, another little boy insisted that Jeremy must be a girl because "only girls wear barrettes." After repeatedly asserting that "wearing barrettes doesn't matter; being a boy means having a penis and testicles," Jeremy finally pulled down his pants as a way of making his point more convincingly. The boy was not impressed. He simply said, "Everybody has a penis; only girls wear barrettes."

Given our Freudian heritage, it would surely be the ultimate irony for developmental psychology if emphasizing the genital difference between the sexes should turn out to be one of the most effective feminist prescriptions for reducing the preschool child's traditional gender traditionalism.

References

- Bem, S. L. (1988). *Rethinking children's thinking about gender: How inevitable is childhood gender traditionalism?* Unpublished manuscript, Cornell University.
- Carey, S. (1985). *Conceptual change in childhood*. Cambridge, MA: MIT/Bradford Press.
- Chi, M. T. H. (1978). Knowledge structure and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.
- DeVries, R. (1969). Constancy of generic identity in the years three to six. *Monographs of the Society for Research in Child Development*, 34(3, Serial No. 127).
- Emmerich, W., Goldman, K. S., Kirsh, B., & Sharbany, R. (1977). Evidence for a transitional phase in the development of gender constancy. *Child Development*, 48, 930–936.
- Fagot, B. I. (1977). Consequences of moderate cross-gender behavior in preschool children. *Child Development*, 48, 902–907.
- Ferguson, G. A. (1981). *Statistical analysis in psychology and education*. New York: McGraw-Hill.
- Flavell, J. H. (1963). *The developmental psychology of Jean Piaget*. New York: Van Nostrand.
- Freud, S. (1925/1959). Some psychological consequences of the anatomical distinction between the sexes. In J. Strachey (Ed.), *Sigmund Freud: Collected papers* (Vol. 5, pp. 186–197). New York: Basic.
- Gelman, R., & Baillargeon, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), P. H. Mussen (Series Ed.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 167–230). New York: Wiley.
- Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number*. Cambridge, MA: Harvard University Press.
- 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.
- Goldman, R. J., & Goldman, J. D. G. (1982). *Children's sexual thinking*. London: Routledge & Kegan Paul.
- Green, B. F. (1956). A method of scalogram analysis using summary statistics. *Psychometrika*, 21, 79–88.
- Keil, F. C. (in press). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT/Bradford Press.
- Kohlberg, L. (1966). A cognitive-developmental analysis of children's sex-role concepts and attitudes. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 82–173). Stanford, CA: Stanford University Press.
- Leinbach, M. D., & Fagot, B. I. (1986). Acquisition of gender labeling: A test for toddlers. *Sex Roles*, 15, 655–656.
- Marcus, D. E., & Overton, W. F. (1978). The development of cognitive gender constancy and sex role preferences. *Child Development*, 49, 434–444.
- Martin, C. L., & Halverson, C. F. (1983). Gender constancy: A methodological and theoretical analysis. *Sex Roles*, 9, 775–790.
- McConaghy, M. J. (1979). Gender permanence and the genital basis of gender: Stages in the development of gender identity. *Child Development*, 50, 1223–1226.
- McConaghy, M. J. (1980). The gender understanding of Swedish children. *Child Psychiatry and Human Development*, 11, 19–32.
- Miller, A. (1984). A transitional phase in gender constancy and its relationship to cognitive level and sex identification. *Child Study Journal*, 13, 259–275.
- Piaget, J. (1932). *The moral judgment of the child*. London: Kegan Paul.
- Piaget, J. (1970). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, pp. 703–732). New York: Wiley.
- Siegel, S. (1956). *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill.
- Slaby, R. G., & Frey, K. S. (1975). Development of gender constancy and selective attention to same-sex models. *Child Development*, 46, 849–856.
- Wehren, A., & DeLisi, R. (1983). The development of gender understanding: Judgments and explanations. *Child Development*, 54, 1568–1578.