Childhood Sex-Typed Behavior and Sexual Orientation: 
A Conceptual Analysis and Quantitative Review

J. Michael Bailey
Northwestern University

Kenneth J. Zucker
Clarke Institute of Psychiatry

This article reviewed research examining the association between childhood sex-typed behavior and sexual orientation. Prospective studies suggest that childhood cross-sex-typed behavior is strongly predictive of adult homosexual orientation for men; analogous studies for women have not been performed. Though methodologically more problematic, retrospective studies are useful in determining how many homosexual individuals displayed cross-sex behavior in childhood. The relatively large body of retrospective studies comparing childhood sex-typed behavior in homosexual and heterosexual men and women was reviewed quantitatively. Effect sizes were large for both men and women, with men's significantly larger. Future research should elaborate the causes of the association between childhood sex-typed behavior and sexual orientation and to identify correlates of within-orientation differences in childhood sex-typed behavior.

Psychosexual differentiation has been a topic of long-standing interest in developmental psychology. Both classical psychoanalytic theory and learning theory were particularly influential in guiding the first wave of empirical research conducted by developmentalists (see, e.g., Mischel, 1966). Sexologists have also made seminal theoretical and empirical contributions. For example, Money (1955) introduced the term gender role to refer to "all those things that a person says or does to disclose himself or herself as having the status of boy or man, girl or woman, respectively. It includes, but is not restricted to, sexuality in the sense of eroticism" (p. 254).

Over the next two decades, gender role was decomposed into three conceptually distinct parts (see, e.g., Fagot & Leinbach, 1985). First, gender identity was distinguished from gender role. For example, Stoller (1964, p. 453) used the slightly different term core gender identity to describe a young child's developing "fundamental sense of belonging to one sex." Cognitive-developmental psychologists (e.g., Kohlberg, 1966) have used the term gender identity to indicate primarily that a child can accurately discriminate male from female individuals and identify correctly his or her own gender—a task considered by some to be the first stage in gender constancy development.

Compared with Money's (1955) original definition, the term gender role is now defined more narrowly. Many scholars have used the term to refer to behaviors, attitudes, and personality traits that a society designates as masculine or feminine, that is, more "appropriate" or typical for the male or female social role (cf. Huston, 1983; Unger, 1979). In young children, the measurement of gender role behavior includes several easily observable phenomena, including affiliative preference for same versus opposite sex peers, interest in rough-and-tumble play, fantasy roles, toy interests, and dress-up play (see Zucker, 1985). In this article, we use the term sex-typed to refer to those behaviors that have been typically studied as markers of childhood gender identity and gender role.

The third, erotic, component of Money's (1955) original definition of gender role has also been operationalized more narrowly, most commonly under the rubric of the term sexual orientation. In contemporary sexology, sexual orientation refers to whether a person is more strongly aroused sexually by members of his or her own sex, the opposite sex, or both sexes (homosexual, heterosexual, and bisexual, respectively).

The behavioral markers of gender identity and gender role emerge early, typically by ages 2-4 years, and become "consolidated" thereafter (e.g., Fagot, 1985; Huston, 1983). In contrast, sexual orientation appears to be more readily assessed after puberty, as a person's sexual interests and desires become more salient (cf. Meyer-Bahlburg, 1980). At least three models have been proposed about the relation between childhood sex-typed behavior and later sexual orientation. One model hypothesizes a developmental sequence in which gender identity develops before gender role, which, in turn, develops before sexual orientation (e.g., Green, 1974, 1987; Meyer-Bahlburg, 1980). In this model, adult sexual orientation is conceptualized as an end state of psychosexual differentiation, analogous, for example, to formal operations as an end state of cognitive development. Another model reverses this developmental sequence, positing that sexual orientation is apparent early enough in development to influence the expression of sex-typed behavior (e.g., Isay, 1989). A third model gives less attention to the temporal sequence between these two variables and instead emphasizes the possibility that sex-typed behavior and sexual orientation are both influenced by the same factors, such as prenatal sex hormones (see Green, 1985; Zucker, 1990, pp. 12-15).
Although some developmentalists have noted the possible relation between sex-typed behavior and sexual orientation (e.g., Brown, 1957, 1958), others have tended to regard the possibility with skepticism. For example, Serbin (1980) commented, "there is no evidence that highly sex-typed children are less likely to become homosexual than children showing less extreme sex-role conformity" (p. 85). Storms (1983) remarked, "Gay men are no less likely to have masculine traits than heterosexual men; and lesbians are no less likely to have feminine traits than heterosexual women" (p. 3). And Kohlberg, Ricks, and Snarey (1984) concluded that "[Existing longitudinal data reveal] little or no correlation between standard measures of [childhood] masculinity-femininity . . . and . . . heterosexuality [in adulthood]" (p. 130).

The primary goal of this article was to review the evidence concerning the possible association between childhood sex-typed behavior and adult sexual orientation. The literature that bears on this question contains two types of studies. Prospective studies have begun with children with atypical patterns of sex-typed behavior and followed them into adolescence or adulthood, when their sexual orientation can be assessed by structured interview. Retrospective studies have asked heterosexual and homosexual participants to recall relevant behaviors and feelings from childhood. There have been far more retrospective than prospective studies. Both types of studies provide unique information, but both also have unique limitations and have been subject to several criticisms. In this article, we summarize and evaluate these criticisms and then provide a meta-analysis of the retrospective studies.

Prospective Studies

Prospective designs avoid some of the most important methodological pitfalls of retrospective research (see, e.g., Yarrow, Campbell, & Burton, 1970), particularly the problem of biased recall. Moreover, they can be a crucial test of the validity of retrospective data (cf. Bowlby, 1969). There are at least two prospective strategies that might yield useful data. One strategy involves the assessment of a large number of unselected children with regard to their sex-typed behavior. As adults, their sexual orientation could be assessed. Although this strategy would be most informative in assessing the association between children with atypical patterns of sex-typed behavior and adult sexual orientation. As adults, their sexual orientation can be assessed by structured interview. Retrospective studies have asked heterosexual and homosexual participants to recall relevant behaviors and feelings from childhood. There have been far more retrospective than prospective studies. Both types of studies provide unique information, but both also have unique limitations and have been subject to several criticisms. In this article, we summarize and evaluate these criticisms and then provide a meta-analysis of the retrospective studies.

One strategy has not been used in any prospective study to date. A second prospective strategy is to ascertain a sample of youngsters believed to be disproportionately likely to develop the outcome of interest (along the lines of studies of children "at risk" for a major mental disorder, such as schizophrenia [e.g., Watt, Anthony, Wynne, & Rolf, 1984]). This design can considerably reduce the number of participants that need to be studied. In studies of psychosexual differentiation, this strategy has involved identifying children (typically boys) who displayed marked patterns of cross-sex behavior. Many of these children would meet the Diagnostic and Statistical Manual of Mental Disorders (4th. ed.; DSM-IV) diagnostic criteria for gender identity disorder (American Psychiatric Association, 1994). These studies examined the relation between childhood cross-gender identification and later psychosexuality, especially transsexualism and homosexuality (see, e.g., Green, 1974).

Green (1974, 1987) has conducted the most comprehensive prospective study of boys with marked patterns of childhood cross-gender behavior. This study contained a sample of 66 feminine and 56 control boys assessed initially at a mean age of 7.1 years (range = 4–12 years). About two thirds of the boys in each group were followed long enough so that their sexual orientation could be assessed in late adolescence (M = 18.9 years; range = 14–24 years). Data from a semistructured clinical interview were used to rate sexual orientation in fantasy and behavior on Kinsey, Pomeroy, and Martin’s (1948, pp. 636–641) 7-point sexual orientation continuum, where 0 = exclusive heterosexuality and 6 = exclusive homosexuality. Depending on the measure (fantasy or behavior), 75%–80% of the previously feminine boys were either bisexual or homosexual (Kinsey ratings of 2–6) at follow-up as compared with 0%–4% of the control boys.

Green’s (1987) results were similar to those of six other follow-up reports of boys who displayed marked cross-gender behavior (Bakwin, 1968; Davenport, 1986; Kosky, 1987; Leboy, 1972; Money & Russo, 1979; Zuger, 1984). As summarized by Zucker (1990), these reports contained 55 boys seen at follow-up, usually in late adolescence or young adulthood (range = 13–36 years [for details, see Zucker, 1985, 1990]). At follow-up, 5 boys were classified as transsexual (all of whom had a homosexual sexual orientation), 21 as homosexual, 15 as heterosexual, and 14 could not be rated with regard to sexual orientation. Excluding these last 14, 26 of 41 boys (63%) had homosexual orientations. Overall, then, there is clear evidence of a relation between patterns of childhood sex-typed behavior and later sexual orientation.

Although these prospective data confirm a link between childhood sex-typed behavior and sexual orientation for boys, several qualifications need to be made. To date, insufficient numbers of girls have been followed prospectively to draw definitive conclusions about the fit with retrospective studies. Second, the prospective studies consisted largely of clinic-referred children who displayed marked patterns of cross-gender behavior, including gender dysphoria. These children might be conceptualized as falling at the extreme end of a continuum of cross-gender identification. The extent of cross-gender behavior recollected by most homosexual adults is probably not as extreme as that displayed by children who meet DSM-IV diagnostic criteria for gender identity disorder (Friedman, 1988). For these reasons, prospective studies are superior to retrospective studies in determining whether, in general, homosexual individuals exhibit atypical sex behavior during childhood.

Retrospective Studies

Since the early 1960s, researchers of diverse theoretical persuasions have collected, by various methods and measures, recall data on components of childhood gender identity and role in men and women with either an exclusive or predominant heterosexual or homosexual sexual orientation. With a few exceptions, this work has been conducted in the United States and Canada. Every study that we located showed, at least on some measures, a significant difference between heterosexual and homosexual adults in their recall of childhood gender identity and
role patterns, with homosexual adults more likely to recall atypical (i.e., statistically uncommon) patterns of childhood sex-typed behavior. Despite the remarkable consistency in findings across studies, the validity of the evidence linking sex-typed behavior and sexual orientation has been questioned. There have been at least three types of criticisms: concerns about sampling techniques, the lack of "perfect" correlations, and threats to the validity of recall data.

**Sampling Methods**

Kohlberg et al. (1984) commented on the use of samples of "homosexuals in trouble (prison or psychotherapy)" (p. 129). Although the heterosexual controls in such studies were also in prison or in psychotherapy, the use of nonrepresentative samples has been recognized as a potential problem (Harry, 1986). But over the years, the empirical studies that have assessed recall data on sex-typed behavior have also sampled many subjects who were neither in prison nor in psychotherapy (see Zucker, 1987). No one has tested the hypothesis that sampling bias affects the differences in recalled childhood sex-typed behavior between heterosexual and homosexual adults.

**Imperfect Correlations**

Carrier (1986) questioned the evidence establishing a relation between sex-typed behavior and sexual orientation because some adults with a homosexual sexual orientation do not recall any cross-gender behavior during childhood and some adults with a heterosexual sexual orientation do. But this merely means that the association between sex-typed behavior and sexual orientation, if it exists, is not a perfect one. Thus, another aim of the present study was to describe quantitatively the strength of the relation between sex-typed behavior and sexual orientation, which can help in interpreting the importance of group differences (cf. Hyde, 1990).

**Validity of Recall Data**

The most common criticism of the retrospective studies concerns memory distortion or selective recall. Ross (1980) advanced a particularly strong version of the retrospective distortion hypothesis: Homosexual adults did not really have cross-gender traits or behaviors in childhood but merely remembered themselves that way because they have internalized societal stereotypes (Ross, 1980; cf. Houlit, 1983/1984; Kite & Deaux, 1987; Risman & Schwartz, 1988). There is no direct empirical support for the retrospective distortion hypothesis. Ross's (1980) study, while often cited as supporting the hypothesis (see, e.g., Houlit, 1983/1984; Ross, 1984), did not show that homosexual adults' recollections were affected by beliefs about homosexuality and gender roles; in fact, it did not even examine sexual orientation differences in childhood cross-gender behavior. The study merely showed that gay men from Sweden were less likely than gay men from Australia (a more conservative culture than Sweden with respect to gender roles) to believe in such an association.

Selective recall of a less complete nature is also possible. Perhaps as a result of social stereotypes, homosexual adults are more likely than heterosexual adults to recall patterns of childhood cross-gender behavior. This could occur if either homosexual adults overreported, or heterosexual adults underreported, atypical sex-typed behavior. For example, homosexual adults familiar with scientific theories about the relation between childhood sex-typed behavior and sexual orientation might provide exaggerated memories of sex-atypical behavior (Ross, Rogers, & McCulloch, 1978). Harry (1984, pp. 120–121), however, found no differences in extent of recalled "gender nonconformity" between gay men and women familiar with scientific writings versus those who were not, suggesting that that particular version of retrospective distortion cannot account for the sexual orientation differences. Similarly, heterosexual adults may be more likely to forget sex-atypical behavior. However, studies of relevant behaviors suggest that most children have little sex-atypical behavior to forget. Unselected samples of children typically show same-sex gender role preferences, and very few persistently wish to be of the other sex (Huston, 1983; Sandberg, Meyer-Bahlburg, Ehhardt, & Yager, 1993; Zucker, 1985).

Differential selective recall has not been empirically demonstrated to magnify sexual orientation differences in childhood memories; however, it remains the most serious concern regarding the validity of the retrospective findings.

Another problem pertains to simple forgetting or childhood amnesia (Pillemer & White, 1989). As a result of an immature memory system (Pillemer & White, 1989), some adults may be unable to retrieve specific memories pertaining to gender behavior from the preschool years, when such behavior first appears (Fagot, 1985; Huston, 1983). Thus, specific memories tapped by retrospective studies are most likely from after the preschool years. Given the stability of children's gender role preferences (Huston, 1983), however, adult retrospections are probably based on an aggregation of multiple displays of relevant behaviors rather than the "one-moment-in-time" events discussed by Pillemer and White (1989, pp. 312–313). In any case, this memory limitation should be unrelated to sexual orientation. Because the recall of childhood events are no doubt imperfect, investigators have attempted to develop methods of checking the accuracy of such data. One strategy has been to use multiple informants (see, e.g., Brewin, Andrews, & Gotlib, 1993). Bailey, Miller, and Willerman (1993) found statistically significant, albeit modest, positive correlations between the recall of childhood gender behavior on the part of young adult homosexual and heterosexual men and women and that of their mothers. Although this finding deserves replication, it provides some evidence for the validity of the participant's own recollections.

In this article, we provide a quantitative review of the retrospective studies. The mean effect size from such a review provides a rough indicator of how large the differences are. Furthermore, there is a sufficiently large number of retrospective studies to allow the examination of factors that might affect the magnitude of effect sizes.

**Study Selection**

Retrospective studies selected for analysis were drawn from the English-language literature. Two criteria were used for inclusion: (a) The
age period for which relevant behaviors were to be recalled was childhood, typically operationalized in relevant measures as either before puberty or ≤12 years of age, and (b) relevant data were available for both a homosexual and a same-sex heterosexual group.

Search procedures included an issue-by-issue search of all volumes of *Archives of Sexual Behavior, Journal of Sex Research,* and *Journal of Homosexuality.* Two bibliographies on research pertaining to homosexuality (Dynes, 1987; Weinberg & Bell, 1972) were also inspected. In addition, we conducted a computer search of the literature using PsycLIT, Medline, and Dissertation Abstracts International (through 1992) using the keywords *homosexuality, gay, lesbian,* and *sexual orientation.* These keywords were then crossed with the keywords *gender identity, gender role,* *sex role,* and *transsexualism.* We also examined articles cited in the bibliographies of relevant studies. Finally, we examined all abstracts of papers presented at the annual meetings of the International Academy of Sex Research. In two instances (Careto, 1991; Guiboin, 1983), we needed to contact authors to obtain data useful for the meta-analyses. Although another study (Tweedy, 1984) collected relevant data, they were no longer available, and so the study was discarded from our analysis. Altogether, there were 41 citations that met the inclusion criteria and that included the necessary data for analysis.

**Coding**

**Study effect sizes.** Effect sizes were computed by using $d$ (Cohen, 1988), calculated as the difference between the means of the homosexual and same-sex heterosexual groups divided by their pooled standard deviations. Because $d$ is a biased estimator, we used Hedges’s (1982, Formula 4) unbiasing correction for all effect sizes. Effect sizes were scaled so that positive values indicate greater recalled cross-gender behavior and identity among the homosexual groups.

Our decision to use the pooled standard deviation requires some justification because, as we show later, the homosexual and heterosexual standard deviations differed from each other. Glass, McGaw, and Smith (1981) have argued that in such cases, one should choose either standard deviation, rather than the pooled standard deviation, because effect sizes computed with the former approach are more meaningful. We rejected that approach for two reasons. First, because (as we also show later) the distributions were not both normal, there can be no straightforward translation from $d$ to a more meaningful description, such as percentage nonoverlap, commonly provided in meta-analyses. Thus, in the present case, $d$ is not very meaningful regardless of the standard deviation used to compute it. Second, because it is computed by using the entire sample, the pooled standard deviation provides the most stable estimate of $d.$ This is important because one primary goal of this meta-analysis is to examine whether various factors (e.g., sex) influence the magnitude of $d.$

Most effect sizes were calculated from means and standard deviations reported in the studies. When this information was unavailable, we used other statistics to estimate the effect size (see Glass et al., 1981, chap. 5).

Some studies reported results for scales of sex-typed behavior. Others reported individual item data, and a few studies reported results for both overall scales and individual items. Because different effect sizes from the same study are not independent, we computed one overall effect size for each study as follows: If results were given for overall scales, these were used because scale scores are more reliable (and hence should yield more stable estimates) than individual items. If only individual item data were given, then effect sizes were computed for all relevant items and then averaged to yield the overall effect size. A total of 48 independent effect sizes were computed for this review’s 41 different citations: 16 for women and 32 for men. Seven citations included relevant data for both men and women. We use the term *study* to refer to each independent effect size. Thus, citations that include data for men and women count as two studies.

**Study characteristics.** We coded several study characteristics to investigate their influence on the magnitude of effect sizes. One factor that we hypothesized would be important was the type of measure used in the study. Some studies reported results on the basis of multi-item scales, whereas others reported results separately for each item examined. Because scale measures are more reliable, they should yield larger effect sizes than individual item measures. Thus, we distinguished studies that used multi-item scales from those that reported data only at the item level.

Sex was also considered, because we hypothesized that effect sizes would be larger for men than for women. Masculine behavior in young girls appears to be more common than feminine behavior in young boys (Huston, 1983).

Several other study characteristics were examined. Year of publication was included to investigate cohort effects. The average age of the sample participants was used to investigate possible age effects on memory. Ascertainment method refers to whether homosexual participants were volunteers from the community or were recruited as clients of mental health professionals. The first author’s discipline from each study was dichotomized into medical versus nonmedical. We analyzed the importance of the first author’s sexual orientation, dichotomized into homosexual versus nonheterosexual. This analysis, requested by a reviewer, is useful in examining the importance of unmeasured biases (e.g., ideology) that might be related to authors’ sexual orientation and influence the magnitude of results. Finally, we coded whether each effect size was computed by using published or unpublished data to investigate whether the unpublished work that we found was unrepresentative. This was particularly important because four of the seven effect sizes from unpublished data were contributed by one of us (JMB).

Table 1 gives frequencies for some relevant predictor variables.

**Table 1**

<table>
<thead>
<tr>
<th>Attribute</th>
<th>No. of studies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures</td>
<td></td>
</tr>
<tr>
<td>Multi-item scale</td>
<td>31</td>
</tr>
<tr>
<td>Single item</td>
<td>17</td>
</tr>
<tr>
<td>Participants</td>
<td></td>
</tr>
<tr>
<td>Women</td>
<td>16</td>
</tr>
<tr>
<td>Men</td>
<td>32</td>
</tr>
<tr>
<td>Homosexual participants ascertained as patients</td>
<td>5</td>
</tr>
<tr>
<td>Homosexual participants ascertained as</td>
<td></td>
</tr>
<tr>
<td>community volunteers</td>
<td>43</td>
</tr>
<tr>
<td>Discipline of lead authors</td>
<td></td>
</tr>
<tr>
<td>Medical</td>
<td>10</td>
</tr>
<tr>
<td>Nonmedical</td>
<td>38</td>
</tr>
<tr>
<td>Sexual orientation of lead authors</td>
<td></td>
</tr>
<tr>
<td>Homosexual</td>
<td>17</td>
</tr>
<tr>
<td>Heterosexual</td>
<td>21</td>
</tr>
<tr>
<td>Unpublished</td>
<td>10</td>
</tr>
<tr>
<td>Unknown</td>
<td>7</td>
</tr>
</tbody>
</table>

Note. A study consisted of an independent effect size. If the same citation included results for both men and women, it counted as two studies. There were a total of 48 studies.

1 For three studies (Bell et al., 1981; Kinsey et al., 1948; Kinsey et al., 1953), effect sizes were computed by the Kinsey Institute using the studies’ raw data.
in more detail. Furthermore, we examined the effect sizes of different types of items.

We coded all of the items according to two separate classifications. The first had seven categories: (a) rough-and-tumble play, competitive athletics, or aggression; (b) toy and activity preference; (c) imagined roles, careers, or role models; (d) cross-dressing; (e) preference for affiliating with male versus female individuals; (f) social reputation as a "sissy," "tomboy," or "loner"; and (g) gender identity. Two raters independently assigned each item to exactly one category, with good interrater reliability ($\kappa = .87$). Only items that both raters assigned to the same categories were coded for later analysis.

The second classification scheme grouped items according to whether they concerned masculine or feminine behavior. For example, the item "I liked wearing dresses" concerns feminine behavior, whereas the item "I liked football" concerns masculine behavior. Some items were neither strictly masculine nor strictly feminine in content (e.g., "I preferred masculine to feminine activities"). Two raters independently rated each item as masculine, feminine, or neither. Interrater reliability was high ($\kappa = .91$). Only items that both raters agreed were either masculine or feminine were coded for later analysis.

Each study contributed at most one observation to each analysis of items. If a study had more than one item in any category, the effect sizes for those items were averaged to give the study's effect size for the category. Thus, each study yielded at most nine item-level effect sizes: one for each of the seven categories from the first classification scheme, a masculine score, and a feminine score.

**Results**

Sample sizes ranged from 34 to 8,751, with a median of 189. The proportion of participants in a study who were homosexual ranged from .025 to .886, with a median of .50. (Both extremes were from large studies, so that even the less frequent group was adequately represented in them.) The total number of participants studied included 8,963 homosexual women, 1,729 lesbians, 11,298 heterosexual men, and 5,734 gay men. After the large samples from the two Kinsey studies (Kinsey et al., 1948; Kinsey, Pomeroy, Martin, & Gebhard, 1953) were excluded, the respective figures were 1,583, 1,539, 3,315, and 4,966. Mean age of the heterosexual subsamples was 27.4 years ($SD = 5.8$); for homosexual subsamples, it was 29.8 years ($SD = 4.7$).

Effect sizes ranged from 0.50 to 2.09, with a mean of 1.19 ($SD = 0.43$), which was significantly greater than zero, $t(47) = 19.2, p < .001$. The mean of effect sizes computed using multi-item measures, 1.37 ($SD = 0.38$), was significantly greater than that for effect sizes computed as the average of single item effects, 0.87 ($SD = 0.32$), $t(46) = 4.7, p < .001$. Type of measure accounted for a substantial proportion of variance in effect size ($R^2 = .32$).

We investigated the influence of the other factors on the magnitude of effect sizes using a regression approach. First, for each factor, we estimated a simple regression model with that factor as the sole predictor of effect size. This allowed us to estimate the importance of each predictor ignoring all other predictors. Second, we estimated a multiple regression model with both the factor of interest and type of measure (i.e., multi-item scales vs. single items) as predictors. Because type of measure was substantially associated with effect size, including it as a covariate should provide a more powerful test of other factors uncorrelated with it. Results from these analyses are provided in Table 2.

Sex was a significant predictor of effect size, both alone, $R^2 = .14$, $F(1, 46) = 7.7, p < .01$, and adjusted for type of measure, $\Delta R^2 = .10, F(1, 45) = 7.5, p < .01$. The mean effect size for men, 1.31 ($SD = 0.43$), exceeded that for women, 0.96 ($SD = 0.35$). Jointly, sex and type of measure accounted for 42% of the total variance in effect sizes. No other predictor was significant either by itself or when type of measure was controlled, though two approached significance. When type of measure was controlled, effect sizes for published studies were larger than those for unpublished studies, $\Delta R^2 = .05, F(1, 45) = 3.6, p = .06$. This may partly reflect the relatively high proportion of unpublished studies of women, which showed smaller effects (57% of unpublished studies vs. 29% of published studies). Again, when type of measure was controlled, effect sizes were larger for studies whose first authors were physicians, $\Delta R^2 = .03, F(1, 45) = 2.3, p = .14$. However, author's discipline was also confounded with sex of subjects, with 80% of physicians' studies about men compared with 63% of studies by authors from nonmedical disciplines.

**Distributions of Sex-Typed Behavior**

Effect sizes are sometimes used to generate indexes of nonoverlap between two distributions (e.g., Smith, Glass, & Miller, 1980), which can be easily obtained by using statistical tables (e.g., Cohen, 1988, p. 22) provided that both distributions are normal and have equal variances. Both assumptions are untenable for studies in the present review. Frequency distributions were available for 12 studies. In every one of these studies, the heterosexual distribution was positively skewed ($p < .001$). The coefficient of skewness for heterosexual distributions ranged from 2.14 to 3.50, with a median of 1.01. Because the heterosexual distributions were positively skewed, they were not normal. In contrast, skewness coefficients for the homosexual distributions ranged from $-4.57$ to $-4.55$, with a median of $-2.00$, and were in all cases less than the heterosexual distributions from the same studies. Thus, compared with the homosexual distributions, the heterosexual distributions were weighted relatively more toward the low end, suggesting a floor effect. This possibility was further supported by an examination of the relative variability of the two groups. Standard deviations were available for 24 studies. In every one of these studies, the ratio of the homosexual to heterosexual standard deviation ($n$) exceeded 1 (median $n = 1.37$), a highly significant excess ($p < .001$).

Because nonnormality precluded computation of indexes of nonoverlap using $Z$ tables, we elected to estimate nonoverlap directly from available data, as follows: Frequency distributions were available for 5 female and 7 male samples given multi-item scales (analyses were restricted to data that used multi-item scales because they should be least affected by measurement error). For each pair of same-sex distributions from the same study, the homosexual and heterosexual subsamples were weighted to equate their sample sizes, and the combined groups' scores were standardized. Equating the subsample sizes maximized the comparability of standard scores across studies because otherwise, the means and standard deviations used to compute standard scores would reflect the ratio of homosexual to heterosexual subjects, which differed between studies. All female samples were weighted to equate their sample sizes, and their distributions were pooled. The male samples were treated similarly and then were pooled to form a separate distribution.
The resulting distributions thus represent an unweighted average of the constituent samples.

Figures 1 and 2 represent the frequency distributions of the pooled samples of women and men, respectively. The effect sizes associated with the two distributions were 1.02 (for women) and 1.53 (for men). These were similar to the effect sizes predicted by regression for studies that used multi-item scales (1.17 and 1.45, respectively); thus, they appear to be fairly representative of the studies examined herein. The female distribution indicated that 81% of lesbians exceeded the heterosexual female median and that 12% of heterosexual women scored exceeded the homosexual female median. There was slightly less overlap for men, reflecting the larger effect size for the male distribution, with 89% of gay men exceeding the heterosexual median, and only 2% of heterosexual men scoring above the homosexual median.

**Item Content**

Table 3 contains the mean effect size for each category of items, separately for men and women. All mean effect sizes were significantly greater than zero (two-tailed tests), with one exception: "Careers/role models" for women, \( t(3) = 2.3, p = .11 \) (\( M = 0.67, SD = 0.59 \)), which was based on only four studies.

---

**Table 2**

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Total effect</th>
<th>Incremental effect</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( R^2 )</td>
<td>( F )</td>
</tr>
<tr>
<td>Type of measure (scale vs. items)</td>
<td>.32</td>
<td>21.8**</td>
</tr>
<tr>
<td>Female vs. male participants</td>
<td>.14</td>
<td>7.7*</td>
</tr>
<tr>
<td>Age of sample</td>
<td>.03</td>
<td>1.1</td>
</tr>
<tr>
<td>Year of publication</td>
<td>.02</td>
<td>1.0</td>
</tr>
<tr>
<td>Patient ascertainment</td>
<td>.00</td>
<td>0.1</td>
</tr>
<tr>
<td>Discipline of first author</td>
<td>.01</td>
<td>0.4</td>
</tr>
<tr>
<td>Sexual orientation of first author</td>
<td>.01</td>
<td>0.4</td>
</tr>
<tr>
<td>Published vs. unpublished</td>
<td>.01</td>
<td>0.6</td>
</tr>
</tbody>
</table>

*Note.* With two exceptions, degrees of freedom are as follows: For total effects, \( F \) values are for 1 and 46 degrees of freedom. Incremental effects represent variance accounted for beyond that explained by type of measure (Predictor 1), and associated \( F \) values are for 1 and 45 degrees of freedom. For age of sample, tests had 1 and 33 and 1 and 32 degrees of freedom, respectively. For sexual orientation of first author, tests had 1 and 36 and 1 and 35 degrees of freedom, respectively. *\( p < .01 \). **\( p < .001 \).
Because the first coding scheme had seven categories, there were too few studies to sustain a statistically powerful test that the categories differed in their average effect sizes. In contrast, it was possible to compare the two categories in the second coding scheme. Specifically, we examined, separately for men and women, whether homosexual and same-sex heterosexual subjects differed more on items assessing masculine behavior or on those focusing on feminine behavior. To test this hypothesis, we performed paired t tests over all studies that had both a feminine and a masculine effect size. The result was not significant for women, t(6) = -0.8, p = .47, but was marginally significant for men, t(12) = -1.8, p = .10, who showed a tendency for larger effect sizes on more masculine-oriented items.

Discussion

As our analyses demonstrated for both men and women, research has firmly established that homosexual subjects recall substantially more cross-sex-typed behavior in childhood than do heterosexual subjects. By Cohen’s (1988) rough criteria, effect sizes were large for both men and women. Indeed, they were among the largest effect sizes ever reported in the realm of sex-dimorphic behaviors. Hyde (1990) summarized the effect sizes from several meta-analytic studies of behaviors often assumed to show significant sex differences: verbal ability (d = -.11), mathematical ability (d = .43), components of spatial ability (ds ranged from .13 to .73), and aggression (d = .50). The effect sizes obtained in the present study exceeded those figures by a wide margin (see also Eaton & Enns, 1986).

Effect sizes were larger for men than for women. Because female homosexuality appears to be less prevalent than male homosexuality (Gebhard, 1972), childhood cross-gender behavior is likely to be less predictive of female than of male homosexuality. To illustrate, assume Gebhard’s (1972) estimates for the prevalence of female and male homosexuality, 1.5% and 4%, respectively. Suppose that a child shows a degree of cross-sex-typed behavior that is typical of those who will become homosexual adults (i.e., at the median of the homosexual distribution). With the use of our results concerning nonoverlap to extrapolate to the general population, 0.75% (.50 * .015) of girls both have the requisite degree of cross-sex-typed behavior and will become lesbians, compared with 11.82% (.12 * .985) of girls who are equally cross-gendered and will become heterosexual women. Thus, only about 6% (.0075 / (.1182 + .0075)) of such girls will become homosexual. Analogous computations suggest that 51% of boys with the requisite degree of cross-sex-typed behavior will become homosexual. Thus, early cross-gender behavior appears to be substantially more predictive of homosexuality in men than in women. Because the expected rate of lesbianism is fairly low, prospective studies of masculine girls may require larger samples than those of feminine boys to achieve adequate statistical power.

These calculations also have relevance to the generality of available prospective studies of boys. The expected rate of adult homosexuality for boys at the homosexual median, 51%, is lower than has been obtained in prospective studies of feminine boys (e.g., Green, 1987; Zuger, 1984), suggesting that boys referred to clinics for cross-sex-typed behavior are more extreme in that respect than are typical prehomosexual boys. On the other hand, the two rates are not so different as to suggest that the findings for the clinic-referred boys are entirely inapplicable to prehomosexual boys in general.

The other factor significantly associated with magnitude of effect size, type of measurement, is less theoretically interesting. Psychometric theory clearly predicts that unreliable items should yield smaller effect sizes than more reliable scales, and this was strongly confirmed. One lesson from this meta-analysis, therefore, is the desirability of using multi-item scales. Even if a researcher wanted to distinguish different aspects of sex-typed behavior (e.g., to see which of them is the best predictor of sexual orientation), the optimal strategy is to use multi-item scales to measure them.

Because a strong empirical link between childhood sex-typed behavior and sexual orientation has been established for men in both prospective and retrospective studies, it is likely to be
genuine. Although it is logically possible that the prospective
findings were due to an extremely unrepresentative group of
boys and the retrospective findings to memory bias, a more par-
simonious interpretation is that both kinds of findings reflect a
rather strong association between childhood sex-typed behavior
and sexual orientation. Retrospective findings for women have
not yet been supported by prospective ones. Although we be-
lieve that prospective studies of masculine girls are desirable,
we know of no strong empirical reason to be skeptical of the
retrospective findings. On the other hand, the possibility re-
mains that retrospective studies exaggerate the true difference
between homosexual and heterosexual participants in child-
hood sex-typed behavior, because of memory bias, and so the
effect sizes we computed should be regarded cautiously.

Historical and Cross-Cultural Limitations

The retrospective studies we reviewed were limited with re-
spect to both time and place. With the exception of Kinsey's
two studies (Kinsey et al., 1948; Kinsey et al., 1953), all of the
studies were published after 1960. On the other hand, the sam-
ple cohorts spanned a period during which important changes
occurred regarding the status and conception of homosexuality.
Some of the homosexual adult samples were from age cohorts
born before the emergence of the modern gay subculture.
Because year of publication was uncorrelated with effect size,
possible--heterosexual differences in childhood sex-typed be-
behavior appear to have been stable across the three decades dur-
ing which the bulk of the studies were conducted.

Almost all of the retrospective studies consisted of homosex-
ual and heterosexual adults from Western industrialized coun-
tries, specifically the United States, Canada, and Australia. Only
Whitam and Mathy (1986, 1991) studied participants from
other countries (Brazil, Guatemala, Peru, and the Philippines).
Although effect sizes for these studies were similar to those con-
taining only Western samples, the generality of the association
between childhood sex-typed behavior and sexual orientation
remains an open question. Whitam's sampling techniques have
been criticized (Risman & Schwartz, 1988), and no one to our
knowledge has attempted to replicate his findings.

Herdt (1990) has described age-structured homosexuality
among Papua New Guinea men in the Sambia tribe. Beginning
in middle childhood, boys are introduced to homosexual activ-
ity (fellatio) with late adolescent and young adult males as a rite
of passage to adult heterosexuality. There is no evidence that
childhood cross-sex-typed behavior is associated with this form
of passage to adult heterosexuality. There is no evidence that
the hypothalamus may affect sexual orientation whereas a sep-
ate area, subject to similar influences during sexual differen-
tiation, affects childhood sex-typed behavior. The plausibility of
this explanation is supported somewhat by findings of LeVay
(1991) and Allen and Gorski (1992), who demonstrated associ-
ations with male sexual orientation for two different areas of the
brain.

The general hypothesis that childhood sex-typed behavior
and sexual orientation are subject to similar hormonal influ-
ences (whether or not they are affected by the same brain struc-
tures) has received empirical support from studies of girls and
women with congenital adrenal hyperplasia (CAH), which re-
results in prenatal and early postnatal exposure to high levels of
androgens. Girls and women with CAH have been found to be
somewhat masculine with regard to several components of sex-
typed behavior (e.g., Berenbaum & Hines, 1992; Ehrhardt
& Baker, 1974). Adult women with CAH also recall masculine
patterns of childhood sex-typed behavior (Dittmann et al.,
1990; Zucker et al., 1992) and appear to have increased rates of
bisexuality and homosexuality, particularly in fantasy (Ditt-
mann, Kappes, & Kappes, 1992; Ehrhardt, Evers, & Money,
1968; Money, Schwartz, & Lewis, 1984; Zucker et al., 1992; but
see Mulaikal, Migeon, & Rock, 1987).

Psychosocial interpretations. Psychosocial interpretations
of the association between childhood sex-typed behavior and
sexual orientation have focused primarily on two interrelated
processes: the influence of parent-child relationships on identi-
fication with parents and differential reinforcement of sex-
typed behavior.

Psychoanalysts have stressed the child's relative identifica-
tion with the same-sex or opposite-sex parent as a precursor to event-
tual sexual orientation (e.g., Bieber et al., 1962). Much of this
literature has focused on men, in whom homosexuality was hy-
pothesized to result from the combination of an excessively
close mother-son relationship and a distant, if not antagonistic,
father-son relationship. This pattern allegedly led to the son's
identification with the mother instead of the father, as was
thought to occur for heterosexual men, and the first manifesta-
tion of this process was hypothesized to be cross-sex-typed be-
havior during childhood. Retrospective studies have generally
found gay men to recall more distance from their fathers and

Nature of the Link Between Childhood Sex-Typed
Behavior and Sexual Orientation

Why is there an association between childhood sex-typed be-
behavior and sexual orientation? Two general approaches to this
question have been explored, one primarily biological and the
other psychosocial.

Biological interpretations. The most prominent biological
hypothesis is that sexual orientation is a function of the degree
of masculinization and defeminization of relevant neural struc-
tures due to prenatal or early postnatal effects of androgens
(e.g., Dörner, 1976; Meyer-Bahlburg, 1984). This theory could
account for the association between childhood sex-typed behav-
or and sexual orientation in at least two ways. First, the relevant
neural structures affecting adult sexual orientation might be the
same structures that, in childhood, affect sex-typed behavior.
The hypothalamus has been the most frequently mentioned
brain area hypothesized to affect sexual orientation. The first
biological explanation for the association between childhood
sex-typed behavior and sexual orientation, then, would also re-
quire that the hypothalamus influence other sex-dimorphic be-
havior. A second possible explanation is that the processes that
masculinize or defeminize the neural structures affecting sexual
orientation also have more general effects. Thus, for example,
the hypothalamus may affect sexual orientation whereas a sepa-
rate area, subject to similar influences during sexual differen-
tiation, affects childhood sex-typed behavior. The plausibility of
this explanation is supported somewhat by findings of LeVay
(1991) and Allen and Gorski (1992), who demonstrated associ-
ations with male sexual orientation for two different areas of the
brain.

Psychosocial interpretations. Psychosocial interpretations
of the association between childhood sex-typed behavior and
sexual orientation have focused primarily on two interrelated
processes: the influence of parent-child relationships on identi-
fication with parents and differential reinforcement of sex-
typed behavior.

Psychoanalysts have stressed the child's relative identifica-
tion with the same-sex or opposite-sex parent as a precursor to event-
tual sexual orientation (e.g., Bieber et al., 1962). Much of this
literature has focused on men, in whom homosexuality was hy-
pothesized to result from the combination of an excessively
close mother-son relationship and a distant, if not antagonistic,
father-son relationship. This pattern allegedly led to the son's
identification with the mother instead of the father, as was
thought to occur for heterosexual men, and the first manifesta-
tion of this process was hypothesized to be cross-sex-typed be-
havior during childhood. Retrospective studies have generally
found gay men to recall more distance from their fathers and
more closeness to their mothers during childhood (for reviews, see Freund & Blanchard, 1983; Friedman, 1988); however, the effect sizes have been modest (e.g., Bell, Weinberg, & Hammer-smith, 1981), and, more important, there is at least one alternative explanation for such findings. Because fathers are relatively intolerant of feminine behavior in their sons (e.g., Langlois & Downs, 1980), they may behave in a more distant or rejecting manner toward prehomosexual sons precisely because those sons are more likely to exhibit such behavior. By this account, childhood cross-sex-typed behavior is a cause rather than a consequence of “father distance.” Available data cannot definitively resolve which, if either, of these two possibilities explains the covariation among childhood sex-typed behavior, father distance, and sexual orientation for men. Biological theories of sexual orientation are more consistent with the possibility that childhood sex-typed behavior and sexual orientation have common influences that precede (and, hence, are more likely to affect than be influenced by) family relationships.

To the extent that biological theories are supported, identification explanations become less plausible. Pillard (1990) provided evidence more directly relevant to this question. Heterosexual brothers of both homosexual and heterosexual index participants rated their closeness to their fathers during childhood. Brothers of homosexual probands rated their fathers as more distant than did the brothers of heterosexual participants, suggesting that the association between sexual orientation and father distance cannot be completely explained by fathers’ reactions to their homosexual sons’ cross-gender behavior. Pillard’s findings suggest that paternal traits may contribute to the emotional distance between gay men and their fathers, but prospective designs would be necessary to resolve the temporal sequence of boys’ cross-gender behavior and fathers’ emotional distance.

The second type of psychosocial explanation of the association between childhood sex-typed behavior and sexual orientation stresses the role of parental socialization. Observational studies of parent–child interactions have provided some evidence of sex-differentiated parental socialization during the toddler and preschool years (e.g., Fagot & Hagan, 1991; Fagot & Leinbach, 1989). Although parental socialization may account for some of the sex differences in childhood sex-typed behavior, these data are not directly related to differences between heterosexual and same-sex homosexual subjects, which are within-sex differences. Although there is some evidence that within-sex variation in the acquisition of sex-typed behavior is associated with parental socialization, this effect appears to diminish by the preschool years (see, e.g., Fagot & Leinbach, 1989). Moreover, the range of behavioral variation is not great and is well within what might be described as the typical range for a child of a given sex. Green (1987) and colleagues (Roberts, Green, Williams, & Goodman, 1987) have examined socialization of sex-typed behavior among parents of feminine boys and reported that such parents were unusually tolerant of such behavior when it first emerged. These data are insufficient, however, to exclude the possibility that parents generally would be more tolerant than expected in the face of persistent cross-sex-typed behavior in their children.

Even if parental socialization accounted for the induction or maintenance of childhood cross-sex-typed behavior, the question would remain how such behavior is associated with, or converted to, a homosexual orientation. Green (1980) and Thorne (1986) suggested that the peer group might be an important context in which stimuli are eroticized. Green (1980), for example, speculated that feminine boys, because of their affiliative preference for girls, are more likely to assimilate the erotic fantasies about other boys common to the sexual rehearsal play of girls. A similar scenario could be constructed concerning masculine girls. Because little systematic empirical research has focused on the socialization of sexuality and eroticism, related hypotheses are difficult to evaluate.

**Within-Orientation Differences in Childhood Sex-Typed Behavior**

Although there were large mean sexual orientation differences in childhood sex-typed behavior for both sexes, there was also overlap. For the composite distributions we constructed, sexual orientation accounted for less than half the variance in childhood sex-typed behavior in both men and women (37% and 21%, respectively). Thus, the bulk of the variance was within groups. Some of the within-group variance was undoubtedly caused by measurement error. Nevertheless, it seems likely that there are true within-group differences in childhood sex-typed behavior for both heterosexual and homosexual participants. Such differences may be particularly pronounced among homosexual participants, who showed greater variability than heterosexual participants on relevant measures.

**Etiological Implications.** Any etiological theory of sexual orientation must account for, or at least allow, within-orientation differences in childhood sex-typed behavior. A neuroendocrine theory of sexual orientation might hypothesize that the brain structures affecting childhood sex-typed behavior and sexual orientation differentiate at similar, but not identical, times. Thus, any important fluctuations in androgens and other relevant hormones would be relatively likely to influence both sex-typed behavior and sexual orientation, but could also affect only one of them if sufficiently delimited.

Goy, Bercovitch, and McBrair (1988) provided some experimental support for this general idea in a study of prenatally androgenized female rhesus macaques. They varied the timing of prenatal exogenous exposure to testosterone propionate and found differential effects on selected sex-dimorphic behaviors. Compared with normal female rhesus macaques, early-exposed female rhesus macaques showed increased rates of maternal and peer-mounting (male-typical behaviors) but did not differ in their rates of rough play; in contrast, late-exposed female rhesus macaques showed increased rates of rough play and peer-mounting but did not differ from normal female rhesus macaques in their rate of maternal-mounting. Goy et al. concluded that “the individual behavior traits that are components of the juvenile male role are independently regulated by the organizing action of androgen and have separable critical periods” (p. 552).

It is also conceivable that neuroendocrine theories might explain only homosexuality in individuals with childhood cross-sex-typed behavior and that different etiological hypotheses are needed to account for homosexual individuals who behaved more typically with respect to such behavior. For example, Bell et al. (1981) speculated that homosexuality not associated with cross-sex-typed behavior was more psychosocially determined.
Because neither specific biological nor psychosocial influences on sexual orientation have been definitively demonstrated, there is no straightforward test of this hypothesis. Indirect tests have been attempted in both men and women with a behavior genetics framework (Bailey & Pillard, 1991; Bailey, Pillard, Neale, & Agey, 1993). The likelihood that homosexual monozygotic twins were from discordant or discordant pairs did not depend on the extent of childhood cross-sex-typed behavior, suggesting that homosexuality associated with childhood cross-sex-typed behavior is not especially genetic compared with other forms of homosexuality. Future etiological studies of sexual orientation should also assess childhood sex-typed behavior and should attempt to discern whether putative etiological factors affect sexual orientation alone or in combination with such behavior. For example, studies of women with CAH might assess both sexual orientation and childhood sex-typed behavior. If CAH is an appropriate model only for lesbians with childhood cross-sex-typed behavior, then lesbians with CAH should have a higher mean and restricted variance on retrospective measures of childhood sex-typed behavior compared with more representative samples of lesbians.

Developmental continuity of childhood sex-typed behavior. A number of studies have assessed gender-related personality traits in homosexual versus heterosexual adults using instruments such as the Bem Sex Role Inventory, the Personality Attributes Questionnaire, and the Masculinity–Femininity subscale of the Minnesota Multiphasic Personality Inventory. Although most of these studies have found homosexual adults to respond somewhat like opposite-sex heterosexuals on these scales (for a review, see Pillard, 1991), they have generally ignored the question of whether adult personality differences are related to childhood differences in sex-typed behavior, that is, whether orientations, the most sex-atypical adults were the most sex-atypical children. Two studies of homosexual adults found some support for continuity (Hooberman, 1979; Saghir & Robins, 1973), but further research on this question is desirable.

Other correlates of childhood sex-typed behavior. Several characteristics have been found to correlate with childhood sex-typed behavior within either heterosexual or nonheterosexual (i.e., homosexual and bisexual) samples. Perhaps most relevant to this review, several studies have explored whether sex-typed behavior predicts gradations of sexual orientation within homosexual or nonheterosexual groups. McConaghy and Silove (1991) found that several items measuring sex-typed behavior correlated with degree of homosexual interest in a group of predominantly heterosexual men. Phillips and Over (1992) found similar results for nonheterosexual men. They reported that for each of 10 items regarding sex-typed behavior, bisexual men were intermediate between heterosexual and gay men (sexual orientations were based on Kinsey ratings). A similar analysis by Bailey (1989), however, failed to find such a relation in nonheterosexual men. Bell et al. (1981, p. 208) found that bisexual women reported less “childhood gender nonconformity” than lesbians. McConaghy and Silove (1991) found no relation between sex-typed behavior and sexual orientation within predominantly heterosexual women.

Results from several studies suggest that, for men, atypical patterns of sex-typed behavior during childhood were associated with negative adult outcomes. In the Fels Research Institute's longitudinal study, Kagan and Moss (1962/1983, pp. 156–167) found that relatively masculine boys were less likely to show avoidance and inhibition regarding heterosexual erotic behavior in adulthood than were relatively unmasculine boys. Aube and Koestner (1992) analyzed data from Sears, Maccoby, and Levin's (1957) longitudinal study and found that men who had been relatively unmasculine boys were less well-adjusted socially and personally than were men who had been more typically masculine boys. Sexual orientation was not examined in either study; thus, it is unclear if these findings would remain if sexual orientation was controlled. Three studies of gay men found a similar link between childhood cross-sex-typed behavior and later problems, including lower self-esteem (Harry, 1983a), higher rates of depression and anxiety (Weinrich, Atkinson, Grant, & The HNRC Group, 1992), and suicidality (Harry, 1983b). None of the studies that included women found similar associations for them (Aube & Koestner, 1992; Harry, 1983b; Kagan & Moss, 1983), consistent with other work suggesting that cross-gender behavior in girls is less likely to have maladaptive social sequelae (Fagot, 1977). It is possible that the more negative outcomes for men reflect the greater social intolerance of cross-gender behavior in boys and in men.

Conclusion

Homosexual individuals recall substantially more childhood cross-sex-typed behavior than do heterosexuals of the same sex. Prospective studies have supported these retrospective findings for men; analogous studies for women remain to be done. Future research should focus on the causes of this association, as well as the causes and consequence of within-orientation variation in sex-typed behavior.

References

References marked with an asterisk indicate studies included in the meta-analysis.


Received April 8, 1993
Revision received December 6, 1993
Accepted December 9, 1993