

## A Follow-Up Study of Girls With Gender Identity Disorder

Kelley D. Drummond

Ontario Institute for Studies in Education of the University of  
Toronto

Susan J. Bradley

Centre for Addiction and Mental Health

Michele Peterson-Badali

Ontario Institute for Studies in Education of the University of  
Toronto

Kenneth J. Zucker

Centre for Addiction and Mental Health

This study provided information on the natural histories of 25 girls with gender identity disorder (GID). Standardized assessment data in childhood (mean age, 8.88 years; range, 3–12 years) and at follow-up (mean age, 23.24 years; range, 15–36 years) were used to evaluate gender identity and sexual orientation. At the assessment in childhood, 60% of the girls met the *Diagnostic and Statistical Manual of Mental Disorders* criteria for GID, and 40% were subthreshold for the diagnosis. At follow-up, 3 participants (12%) were judged to have GID or gender dysphoria. Regarding sexual orientation, 8 participants (32%) were classified as bisexual/homosexual in fantasy, and 6 (24%) were classified as bisexual/homosexual in behavior. The remaining participants were classified as either heterosexual or asexual. The rates of GID persistence and bisexual/homosexual sexual orientation were substantially higher than base rates in the general female population derived from epidemiological or survey studies. There was some evidence of a “dosage” effect, with girls who were more cross-sex typed in their childhood behavior more likely to be gender dysphoric at follow-up and more likely to have been classified as bisexual/homosexual in behavior (but not in fantasy).

*Keywords:* gender identity disorder, gender identity, sexual orientation, girls, follow-up

Research on normative (or typical) gender development has documented various behavioral domains in which children show, on average, significant sex differences: gender identity self-labeling, sex-of-playmate preference, toy and activity interests, roles in fantasy play, parental rehearsal play, and so on (for a review, see Ruble, Martin, & Berenbaum, 2006; Zucker, 2005c). The determinants of this between-sex variation in sex-typed behavior have long been deemed by developmentalists to have important implications for other aspects of psychosocial development, such as interpersonal relational styles (e.g., Maccoby, 1998), cognitive skills (e.g., Liss, 1983), and vocational interests (e.g., Lippa, 1998), for which there are also significant sex differences.

As noted by Lippa (2002), determining within-sex individual differences in gender-related behavior is another strategy used to study variations with regard to other aspects of development (see, e.g., Barrett & White, 2002; Khuri & Ruble, 2006). In the present study, we used this approach to examine the relation, if any, between sex-typed behavior patterns in childhood, including gen-

der identity, and subsequent gender identity and sexual orientation in late adolescent girls and young adult women.

Several lines of evidence suggest that there are empirical reasons to posit a link between sex-typed behavior in childhood and later gender identity and sexual orientation. Like sex-typed behavior in childhood, gender identity and sexual orientation in adulthood are also sex dimorphic: Most women have a “female” gender identity (the subjective sense of self as a woman) and are sexually attracted to men, whereas most men have a “male” gender identity and are sexually attracted to women. Indeed, gender identity and sexual orientation may be the two behavioral traits that most strongly differentiate women from men (cf. Hyde, 2005). Using a self-report questionnaire designed to measure gender identity dimensionally in adolescents and adults, for example, Deogracias et al. (2007) obtained a between-sex effect size, using Cohen’s *d*, of 13.24.

Over the past several decades, the empirical literature has relied on two methods, namely, retrospective and prospective designs using targeted samples, to examine the relation between sex-typed behavior in childhood and subsequent gender identity and sexual orientation in adulthood. Retrospective designs have studied adults with known variation in their gender identity and/or sexual orientation. For example, adults who meet the *Diagnostic and Statistical Manual of Mental Disorders (DSM)* criteria for gender identity disorder (GID; also known as transsexualism) recall engaging in more cross-gender-typed behavior in childhood than do adults without GID (e.g., Blanchard & Freund, 1983; Doorn, Poortinga, & Verschoor, 1994; Ehrhardt, Grisanti, & McCauley, 1979; Freund, Langevin, Satterberg, & Steiner, 1977; see also Bartlett & Vasey, 2006).

---

Kelley D. Drummond and Michele Peterson-Badali, Department of Human Development and Applied Psychology, Ontario Institute for Studies in Education of the University of Toronto, Toronto, Ontario, Canada; Susan J. Bradley and Kenneth J. Zucker, Gender Identity Service, Child, Youth, and Family Program, Centre for Addiction and Mental Health, Toronto, Ontario, Canada.

Correspondence concerning this article should be addressed to Kenneth J. Zucker, Gender Identity Service, Child, Youth, and Family Program, Centre for Addiction and Mental Health, 250 College Street, Toronto, Ontario M5T 1R8, Canada. E-mail: ken\_zucker@camh.net

The largest body of retrospective research pertains to the within-sex association between sex-typed behavior in childhood and sexual orientation in adulthood. Bailey and Zucker (1995) performed a meta-analysis of 41 retrospective studies that made a quantitative comparison between heterosexual and homosexual same-sex adults using some measure of childhood sex-typed behavior. These studies yielded 48 independent effect sizes: 32 compared heterosexual and homosexual men, and 16 compared heterosexual and homosexual women. Using Cohen's *d*, Bailey and Zucker found that there were substantial differences in patterns of recalled childhood sex-typed behavior between heterosexual and homosexual adults. On average, both homosexual men and women recalled more cross-sex-typed behavior in childhood than did their heterosexual counterparts (respective *d*s were 1.31 and 0.96). Subsequent studies have, with no exception, continued to replicate these findings (summarized in Zucker et al., 2006).

There are, of course, both methodological and interpretive problems with retrospective designs (for an overview, see Hardt & Rutter, 2004). In a targeted sample of adults with GID (invariably recruited from specialized gender identity clinics), it is possible that the association with cross-sex-typed behavior is magnified because not all individuals with pervasive cross-gender behavior in childhood end up seeking out medically assisted gender change in adulthood (e.g., because their earlier gender dysphoria had desisted). In the studies comparing the recollections of heterosexual and homosexual adults, in which there is less of a sampling bias problem, the most common criticism has pertained to memory distortion or selective recall. For example, it has been argued that the greater recollection of cross-gender behavior during childhood by homosexual than by heterosexual adults is linked to the widespread "master narrative" in Western culture that presupposes that "gender inversion" is linked to homosexual sexual orientation (see, e.g., Cohler & Galatzer-Levy, 2000; Gottschalk, 2003; Hegarty, 1999; Kite & Deaux, 1987). As a result, it has been claimed that the sex-typed behavior–sexual orientation association is nothing more than participants recalling behaviors that adhere to cultural stereotypes and expectations. Although there is evidence that speaks against this retrospective distortion hypothesis (summarized in Bailey & Zucker, 1995; Zucker, 2005a, in press; Zucker et al., 2006), there is general agreement that the retrospective data should be confirmed (or disconfirmed) with prospective designs.

One prospective approach has been to target a sample of children presumed to have moderate-to-pervasive cross-gender behavior. In one line of research, sampling consisted of ad-recruited girls with parent-nominated "tomboyish" behavior, along with measures of sex-typed behavior administered to the girls themselves (e.g., Bailey, Bechtold, & Berenbaum, 2002; Berenbaum & Bailey, 2003; Green, Williams, & Goodman, 1982), who were compared to girls unselected for their gender behavior. Neither research team has, as of yet, reported on longer term linkages.

A second strategy has been to study children referred to specialized gender identity clinics because there is concern about their cross-gender behavior and gender identity status (e.g., on the part of parents, mental health professionals, teachers, etc.). Over the years, several research teams have studied such children, and overviews may be found in the work of Green (1987), Zucker and Bradley (1995), and Cohen-Kettenis and Pfäfflin (2003).

In one study, Green (1987) assessed the gender identity and sexual orientation of 44 behaviorally feminine boys and 30 control

boys who were at a follow-up mean age of 18.9 years (range, 14–24 years) and who had initially been evaluated at a mean age of 7.1 years (range, 4–12 years). Of the 44 behaviorally feminine boys, only 1 youth, at the age of 18 years, was gender dysphoric to the extent of considering sex-reassignment surgery. None of the other boys were reported to have gender identity problems at follow-up. Sexual orientation in fantasy and behavior was assessed by means of a semistructured, face-to-face interview. Kinsey ratings were made on a 7-point continuum, ranging from exclusive heterosexuality (a Kinsey "0") to exclusive homosexuality (a Kinsey "6"; Kinsey, Pomeroy, & Martin, 1948). Depending on the measure (fantasy or behavior), 75%–80% of the previously behaviorally feminine boys were either bisexual or homosexual (Kinsey ratings between 2 and 6) at follow-up versus 0%–4% of the control boys.

Data from seven other follow-up reports on a total of 82 behaviorally feminine boys have been summarized in detail elsewhere (Zucker, 2005b; Zucker & Bradley, 1995, pp. 285–286, 290–297). Similar to Green's (1987) case-control study, these studies also identified an elevated rate of either a bisexual or homosexual sexual orientation (52.4%). In contrast to Green's (1987) study, however, the other studies found the rate of GID persistence was higher, with rates ranging from 12% to 20%.

From these prospective studies of behaviorally feminine boys, two conclusions might be drawn: (a) The rate of persistent gender dysphoria was modest but arguably higher than one estimated base rate for gender dysphoria in the general population of biological males: 1 in 11,000 men (Bakker, van Kesteren, Gooren, & Bezeemer, 1993), and (b) the rate of a later bisexual or homosexual sexual orientation was notably higher than the known base rates for a bisexual or homosexual sexual orientation in the general population of biological males (see, e.g., Laumann, Gagnon, Michael, & Michaels, 1994). Thus, for sexual orientation, there appears to be a reasonable convergence between prospective and retrospective studies but, for gender identity, there is more divergence: Many boys with pervasive cross-gender behavior and co-occurring gender dysphoria do not show persistent gender dysphoria by late adolescence or young adulthood, which is at some variance from the recollections of most gender-dysphoric adolescent boys and adult men.

Over the years, it has been noted that little is known about the longer term psychosexual outcome of girls referred to specialized gender identity clinics (Peplau & Huppel, in press; Peplau, Spalding, Conley, & Veniegas, 1999). In part, this has been a function of the fact that boys are much more likely than girls to be referred to gender identity clinics: 5.75:1 in one clinic and 3.07:1 in another (see Cohen-Kettenis, Owen, Kaijser, Bradley, & Zucker, 2003; Cohen-Kettenis et al., 2006). The present study attempted to fill this gap by providing, to our knowledge, the first systematic follow-up report of clinic-referred girls with GID with regard to gender identity and sexual orientation.

## Method

### *Participants*

Between 1975 and 2004, 71 girls (age range, 3–12 years) were referred for assessment to the Gender Identity Service, Child, Youth, and Family Program at the Centre for Addiction and

Mental Health in Toronto, Ontario, Canada. To participate in the follow-up study, patients had to be at least 17 years of age. Using this age cutoff, we identified 37 eligible girls, of whom 30 were contacted for participation. Of the remaining 7 girls, 3 could not be traced through previous addresses, registrars, and personal contacts (e.g., the patient and/or their family had moved and a current telephone number, mailing address, or e-mail address could not be identified), and 4 were not available to participate within the time requirements of the study.

Initial telephone contact was first made with the parents or legal guardians because participants were minors at the time of assessment and some may have had no recollection of their clinic attendance.<sup>1</sup> Of the 30 clients contacted, 25 (83.3%) agreed to participate; 24 came into the clinic for testing, and 1 participant completed a telephone interview because she was too anxious to travel to the clinic. Of the remaining 5 girls, 4 of the girls' parents or guardians (e.g., the Children's Aid Society) were unwilling to provide contact information for their children. One individual declined to participate.

The demographic characteristics of the participants in childhood and at follow-up are shown in Table 1. The GID diagnosis in childhood was based on the *DSM* (3rd ed. [*DSM-III*]; 3rd ed., rev. [*DSM-III-R*]; or 4th ed. [*DSM-IV*]; American Psychiatric Association [APA], 1980, 1987, and 1994, respectively) criteria applicable at the time of assessment. Fifteen girls (60%) met complete

*DSM* criteria for GID in childhood. The remaining 40% were subthreshold for a *DSM* diagnosis of GID, but all had some indicators of GID, and some would have met the complete *DSM* criteria at some point in their lives prior to their assessment in childhood.

Four of the girls in the follow-up sample were born with a disorder of sex development (DSD; 2 had cloacal exstrophy, 1 had congenital micropenis syndrome of unknown etiology, and 1 had mixed gonadal dysgenesis; Hughes, Houk, Ahmed, Lee, & Lawson Wilkins Pediatric Endocrine Society/European Society for Paediatric Endocrinology Consensus Group, 2006). Three of the nonparticipants also had a DSD (partial androgen insensitivity syndrome, congenital adrenal hyperplasia, or true hermaphroditism). There are arguments for and against the inclusion of the 4 girls with a DSD in this sample (see, e.g., Meyer-Bahlburg, 1994). A female gender assignment was made for all 4 girls almost immediately after birth. Also in early infancy, the 4 girls were gonadectomized and had surgical feminization of their external genitalia. Like the somatically intact girls, the 4 girls were referred for concern about their gender development in relation to their assigned gender. On the one hand, as noted by Meyer-Bahlburg (2005), "there is every reason to assume that the processes and psychosocial factors involved in normative gender development also contribute to development of all aspects of gender. . . in persons with intersexuality" (p. 434). On the other hand, as also noted by Meyer-Bahlburg (2005), "additional factors. . . may come into play in [such persons]. . . particularly the awareness of an atypical biological condition and medical history" (pp. 434–435). As noted in Table 3, only 1 of these girls met the complete Point A and Point B *DSM* criteria for GID, and the other 3 were subthreshold.

Table 1  
Demographic Characteristics (*N* = 25)

Characteristic	<i>M</i>	<i>SD</i>	Range	%
From childhood				
Age (in years)	8.88	3.10	3.17–12.95	
Year of assessment	1989.36	7.02	1977–2002	
IQ <sup>a, b</sup>	105.17	21.73	57–144	
Social class <sup>c</sup>	35.72	14.40	8–66	
Marital status <sup>d</sup>				
Two-parent family				60.0
Other				40.0
Caucasian				80.0
At follow-up				
Age (in years) <sup>e</sup>	23.24	5.82	15.44–36.58	
Year of birth	1980.52	6.06	1968–1989	
Interval (in years) <sup>f</sup>	14.34	7.03	2.99–27.12	
IQ <sup>b, g</sup>	10.20	2.71	5.00–15.75	

<sup>a</sup> Full-scale IQ was obtained with age-appropriate Wechsler intelligence scales (the Wechsler Preschool and Primary Scale of Intelligence—Third Edition [Wechsler, 2002], the Wechsler Intelligence Scale for Children—Revised [Wechsler, 1974], and the Wechsler Intelligence Scale for Children—Third Edition [Wechsler, 1991]). One participant was administered the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986). <sup>b</sup> IQ scores at assessment and follow-up were not available for 1 participant. <sup>c</sup> For social class, Hollingshead's (1975) Four Factor Index of Social Status was used. The absolute range was 8–66. <sup>d</sup> For marital status, the category "Other" included the following family constellations: single parent, separated, divorced, living with relatives, or in the care of the Children's Aid Society. <sup>e</sup> One participant (who was 15.44 years of age) was below the lower bound age cutoff of 17 years but was included in the study because her guardian had contacted the clinic for issues unrelated to gender identity status. <sup>f</sup> Interval denotes the time between childhood assessment and follow-up assessment. <sup>g</sup> Composite IQ = (Vocabulary + Comprehension + Block Design + Object Assembly subscale scores)/4. The absolute range was 1–19.

### Procedure

All participants were evaluated on a single day. Below, we provide information on the measures used in this report (for other measures, including parent and self-ratings of behavior problems, psychiatric diagnoses, and experiences of stigma, see Drummond, 2006). All of the participants provided written informed consent prior to their involvement in the follow-up assessment and were provided a stipend for their participation and reimbursement for travel expenses. The study was approved by the Institutional Review Boards at the Centre for Addiction and Mental Health and the University of Toronto.

<sup>1</sup> It is beyond the scope of this report to describe the types of therapies (as well as their frequency and duration) that the girls and/or their parents may have received between the assessment in childhood and the follow-up (e.g., by a therapist within the Gender Identity Service at the Centre for Addiction and Mental Health or in the community). From the participants' clinic files, 13 of the 25 girls had at least some contact with our clinic during the interval between assessment and follow-up (e.g., as therapy clients or for a reassessment). Of the 25 girls and/or their parents, 18 had been in some type of therapy or counseling during the interval between assessment and follow-up; of these, 5 were patients of staff within the Gender Identity Service, and the remainder were seen by a professional in the community.

## Measures

### Childhood Assessment

**Cognitive functioning.** IQ was assessed with the Wechsler Adult Intelligence Scale—Third Edition (Wechsler, 1997) or the Wechsler Intelligence Scale for Children—Third Edition (Wechsler, 1991) and, for one participant, with the Stanford-Binet Intelligence Scale (Thorndike, Hagen, & Sattler, 1986).

**Sex-typed behavior.** Five child informant and three parent informant measures were used to assess the participants' sex-typed behavior in childhood: (a) the Draw-a-Person test (Zucker, Finegan, Doering, & Bradley, 1983); (b) a free-play task (Zucker, Doering, Bradley, & Finegan, 1982); (c) the Playmate and Play-style Preferences Structured Interview (Fridell, Owen-Anderson, Johnson, Bradley, & Zucker, 2006); (d) sex-typed responses on the Rorschach test (Zucker, Lozinski, Bradley, & Doering, 1992); (e) the Gender Identity Interview (Zucker et al., 1993); (f) the Gender Identity Questionnaire for Children (Johnson et al., 2004); (g) a measure of activity level/extraversion (Zucker & Bradley, 1995); and (h) the Games Inventory (Bates & Bentler, 1973). These child and parent informant measures all had established discriminant validity, that is, they significantly differentiated the clinic girls referred for gender identity concerns from control girls (for a review, see Zucker, 2005c; Zucker & Bradley, 1995). A Childhood Sex-Typed Behavior Composite was computed for each participant by averaging the z-scores for these measures (which yielded a total of 11 indices), as well as the GID *DSM* diagnosis (1 = *threshold*, 2 = *subthreshold*) in childhood. Data from the total sample of participants and nonparticipants ( $N = 37$ ) were used. Because of missing data, the mean number of indices/participant was 9.16 ( $SD = 2.30$ ).

### Follow-Up Assessment

**Cognitive functioning.** Four subtests (Vocabulary, Comprehension, Block Design, and Object Assembly) of the Wechsler Adult Intelligence Scale—Third Edition or the Wechsler Intelligence Scale for Children—Third Edition were administered. The standard scores from the subtests were averaged to form an IQ score for cognitive functioning.

**Recalled childhood gender identity and gender role behavior.** Participants completed the Recalled Childhood Gender Identity/Gender Role Questionnaire (RCGI; Zucker et al., 2006). This questionnaire consists of 23 items pertaining to various aspects of sex-typed behavior, as well as to the relative closeness to the mother and father during childhood. Individual items were rated on a 5-point response scale. Each participant was instructed to make ratings for her behavior as a child ("between the years 0 to 12"). Factor analysis identified two factors, accounting for 37.4% and 7.8% of the variance, respectively (all factor loadings  $\geq .40$ ). Factor 1 consisted of 18 items that pertained to childhood gender role and gender identity, and Factor 2 consisted of three items that pertained to parent-child relations (relative closeness to one's mother versus one's father). Information on normative sex differences and discriminant validity was reported in Zucker et al. (2006). For the present study, the mean Factor 1 score was computed for each participant.

**Concurrent gender identity.** During an audiotaped interview, each participant was asked to describe her current feelings about

being female and then to describe positive and negative aspects about her gender status. The examiner also asked semistructured gender identity questions from the adolescent and adult GID criteria outlined in the *DSM-IV-TR* (APA, 2000). The interviewer asked four questions related to the Point A criteria (e.g., the stated desire to be a man, the desire to live or to be treated as a man) and six questions from the Point B criteria (e.g., a preoccupation with getting rid of breasts or genitalia). Participants were asked to respond according to the last 12 months with *No*, *Yes*, or *Sometimes*. Participants who answered *Yes* or *Sometimes* for one or more of the questions from both Point A and B criteria were classified as displaying persistent gender dysphoria.

The female version of the Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults (GIDQ-AA; Deogracias et al., 2007) was also completed. This 27-item questionnaire measures gender identity and gender dysphoria in adolescents or adults. Item content was based on prior measures, expert panels, and clinical experience. Each item was rated on a 5-point response scale ranging from *Never* to *Always* based on a time frame of the past 12 months. Item examples include the following: "In the past 12 months, have you felt unhappy about being a woman?" and "In the past 12 months, have you wished to have an operation to change your body into a man's (e.g., to have your breasts removed or to have a penis made)?" Factor analysis identified a strong one-factor solution that accounted for 61.3% of the variance. All 27 items had factor loadings  $\geq .30$  (median, .86; range, .34–.96). Psychometric evidence for discriminant validity and clinical utility can be found in Deogracias et al. (2007). Participants' GIDQ-AA total scores were calculated by summing scores on the completed items and dividing by the number of marked responses.

**Sexual orientation in fantasy.** Each participant's sexual orientation in fantasy was assessed with specific questions during an audiotaped face-to-face interview and the self-report Erotic Response and Orientation Scale (EROS; Storms, 1980). Questions posed in the interview addressed four types of sexual fantasy: (a) crushes on other people, (b) sexual arousal to visual stimuli (e.g., to strangers, acquaintances, partners, and individuals presented in the media [video, movies, magazines, the internet]), (c) sexual content of night dreams, and (d) sexual content of masturbation fantasies. Using the Kinsey scale criteria, the interviewer assigned ratings that ranged from 0 (*exclusively heterosexual*) to 6 (*exclusively homosexual*) for each parameter. A dummy score of 7 denoted that the participant did not experience or report any fantasies. A global fantasy score was derived on the basis of ratings from the four questions. In the present study, only ratings for the last 12 months are reported.

During the interview, participants were not asked directly about the gender of the person or persons who elicited sexual arousal, thus allowing time for the participant to provide this information spontaneously. Directed questions were asked only if the participant did not volunteer specific information about same-sex or opposite-sex partners. This approach was used so that, by the end of the interview, the participant provided information about sexual arousal to both same-sex and opposite-sex individuals.

The EROS is a 16-item self-report measure assessing sexual orientation in fantasy over the past 12 months. Half of the questions pertained to heterosexual fantasy (e.g., "How often have you had any sexual feelings (even the slightest) while looking at a man?") and the other half pertained to homosexual fantasy (e.g.,



“How often have you had any sexual feelings (even the slightest while looking at a woman?”). Each item was rated on a 5-point scale for frequency of occurrence, ranging from “none” to “almost every day.” Mean homoerotic and heteroerotic fantasy scores were derived for each participant. Previous use of the EROS has shown good evidence of discriminant validity (Storms, 1980; Zucker et al., 1996).

*Sexual orientation in behavior.* Each participant’s sexual orientation in behavior was assessed with specific questions during the face-to-face interview and with a modified version of the Sexual History Questionnaire (SHQ; Langevin, 1985). In the interview, questions asked about five types of sexual behavior: (a) dating; (b) holding hands in a romantic manner; (c) kissing; (d) genital fondling or being touched on the breasts (or, in cases of same-sex sexual behavior, touching another woman’s breasts); and (e) penile–vaginal intercourse, anal intercourse, or the use of dildos. Kinsey ratings for behavior for the past 12 months were made in the same manner as fantasy ratings.

The modified SHQ consisted of 20 questions. Ten questions pertained to heterosexual experiences (e.g., “How many men have you kissed on the lips in a romantic way?”), and 10 questions pertained to homosexual experiences (e.g., “How many women have you kissed on the lips in a romantic way?”). Each item was rated, for the 12 month period prior to the follow-up assessment, on a 5-point scale for frequency of occurrence, ranging from none to 11 or more. Mean total scores for heterosexual and homosexual experiences were derived.

*Sexual identity self-labeling.* Participants were asked to provide a label for their current sexual identity and were offered the following options: (a) “straight” or “heterosexual”; (b) “lesbian,” “homosexual,” or “queer”; (c) “bisexual”; (d) “asexual”; or (e) “other.”

*Social desirability.* Social desirability can threaten the validity of self-report scales when respondents seek social approval or try to represent themselves in a favorable manner (King & Brunner, 2000). Participants  $\geq 18$  years of age completed the Marlowe–Crowne Social Desirability Scale (M–C SDS; Crowne & Marlowe, 1960), which consists of 33 true–false items. The scale consists of 18 culturally acceptable but unlikely statements keyed in the true direction and 15 socially undesirable but probable statements keyed in the false direction for a maximum possible score of 33. Participants under 18 years of age completed a shorter version of the M–C SDS (Strahan & Gerbasi, 1972). This scale consists of 12 culturally acceptable but improbable statements keyed in the true direction and 8 socially undesirable but probable statements keyed in the false direction for a maximum possible score of 20. Several studies have found that the M–C SDS is a reliable and valid measure (Crowne & Marlowe, 1960; Holden & Fekken, 1989; Silverthorn & Gekoski, 1995).

## Results

### *Participants Versus Nonparticipants*

A preliminary analysis compared the assessment information from childhood of the 25 girls who participated in the study with that of the 12 girls who did not participate. There were no significant differences between the participants and nonparticipants on any of these variables (data not shown).<sup>2</sup> At least by these mea-

Table 2  
*Mean Factor 1 Score on the Recalled Childhood Gender Identity/Gender Role Questionnaire (Zucker et al., 2006)*

Group	<i>M</i>	<i>SD</i>	<i>d</i>	<i>n</i>
Total sample	2.57	.67		20
(Female university students)	(3.43)	(.54)		(100)
(Mothers of boys with GID)	(3.80)	(.54)		(230)
(Mothers of control boys)	(3.72)	(.34)		(13)
(Mothers of nonreferred boys)	(3.77)	(.39)		(24)
(Sisters/female cousins of women with CAH)	(3.70)	(.43)		(15)
Childhood diagnosis				
GID: Threshold	2.48	.66	.32	11
GID: Subthreshold	2.70	.69		9

*Note.* Absolute range is 1.00–5.00. A lower score indicates more recalled atypical gender identity and gender role behavior. Groups and values in parentheses are from Zucker et al. (2006). GID = gender identity disorder; CAH = congenital adrenal hyperplasia.

asures, it appears that the participants were representative of the total pool of available patients and thus did not constitute a markedly biased sample at follow-up.

### *Sex-Typed Behavior in Childhood*

Table 2 shows the mean RCGI Factor 1 score, which pertained to the participants’ recollections of their sex-typed behavior from childhood. This mean score can be compared with the scores of several samples of women, unselected for their gender identity or sexual orientation, reported on in Zucker et al. (2006) and also shown in Table 2. By comparing the mean factor score with the scores from the other samples (mean range, 3.43–3.80), we see it is apparent that the women in this study recalled relatively more cross-gender behavior in childhood ( $M = 2.57$ ,  $SD = .67$ ).

Table 2 also shows the mean RCGI Factor 1 score of the participants as a function of *DSM* diagnostic status in childhood. Although the threshold participants recalled, on average, more cross-gender behavior in childhood than the subthreshold participants, the difference was not significant,  $t(18) < 1$ ; the effect size (Cohen’s  $d$ ) of .32 would be considered small. We also examined the  $z$ -composite for childhood sex-typed behavior as a function of diagnostic status (for this analysis, the *DSM* metric was removed from the composite and served as the independent variable). With age at assessment in childhood covaried, the threshold participants had, on average, significantly more cross-sex-typed behavior in childhood ( $M = .15$ ,  $SD = .54$ ) than did the subthreshold participants ( $M = -.31$ ,  $SD = .36$ ),  $F(1, 21) = 23.36$ ,  $p < .001$ , partial  $\eta^2 = .53$ .

### *Psychosexual Differentiation at Follow-Up*

A summary of the psychosexual differentiation data, including gender identity at follow-up, sexual orientation, and sexual identity self-labeling for each participant, is shown in Table 3.

<sup>2</sup> These data are available in the study by Drummond (2006).

Table 3  
Summary of Gender Identity and Sexual Orientation Results at Follow-Up

Participant ID	Age at assessment (years)	Age at follow-up (years)	Global Kinsey ratings		Sexual identity label	Gender identity	DSM
			Fantasy	Behavior			
1	9.74	36.58	6	6	HS	WNL	+
2	8.88	36.61	6	6	HS	WNL	+
3	5.85	32.41	0	0	HT	WNL	+
4	3.17	28.78	0	—	HT	WNL	+
5	4.92	26.61	4	0	BS	WNL	+
6 <sup>a</sup>	5.75	26.58	0	0	HT	WNL	+
7	12.67	17.09	—	—	AS	Dysphoric	
8	12.95	28.72	6	—	HS	WNL	
9	8.41	23.34	6	6	HT	Dysphoric	+
10	8.29	24.12	4	6	BS	WNL	+
11	4.10	20.04	0	0	HT	WNL	+
12	4.72	19.73	0	—	HT	WNL	+
13	6.70	21.53	0	0	HT	WNL	+
14	6.81	18.73	0	0	HT	WNL	+
15	12.62	23.57	6	6	HS	WNL	
16	12.16	21.10	6	6	HT	Dysphoric	+
17	7.32	17.51	0	0	HT	WNL	+
18	8.51	17.34	0	—	HT	WNL	
19 <sup>a</sup>	12.88	21.58	0	—	HT	WNL	
20	9.20	17.81	0	0	HT	WNL	
21	11.26	19.27	0	0	HT	WNL	+
22 <sup>a</sup>	12.18	17.35	—	—	HT	WNL	
23	12.45	15.44	0	—	HT	WNL	
24 <sup>a</sup>	11.89	27.74	0	0	HT	WNL	
25	8.79	23.12	0	0	HT	WNL	

Note. For Kinsey ratings (last 12 months), 0 = exclusively heterosexual and 6 = exclusively homosexual. In the DSM column, a plus sign indicates the participant met complete DSM-III, DSM-III-R, or DSM-IV symptom criteria for gender identity disorder at initial assessment. Dashes indicate the participant did not report fantasy or behavior. ID = identification label; HS = homosexual (lesbian); HT = heterosexual or straight; BS = bisexual; AS = asexual; WNL = within normal limits (i.e., the participant did not report any distress about being a female).

<sup>a</sup> Participant with a disorder of sex development.

### Gender Identity at Follow-up

On the basis of their answers to the semistructured clinical interview questions, participants were classified as either gender dysphoric or not gender dysphoric. In answering these questions, 22 participants (88%) reported no distress with their female gender identity at follow-up. None of the participants desired contrasex hormones or sex reassignment surgery to masculinize their bodies, nor did they express a desire to get rid of their female sex characteristics.

The remaining 3 participants (12%) were classified as gender dysphoric at follow-up (none of these 3 girls had a co-occurring DSD). Among these 3 participants, 1 had been living as a boy since early adolescence (i.e., was known to others as a boy) and was in the process of legally changing his name on official documents. The other 2 participants were living as girls, although both were often perceived of as boys by naïve others (e.g., new acquaintances, strangers, etc.), which they preferred. All 3 gender dysphoric participants wished they had been born a boy and wondered whether they would have been happier as a boy. Two of these individuals indicated a desire to have surgery to masculinize their bodies. The other participant classified as gender dysphoric reported indifference with regard to altering her physical appearance but felt that “it was better to be neutral.” On the basis of this information, 2 of the participants met DSM-IV-TR criteria for GID. Although the other participant did not meet full criteria for

GID, information from the clinical interview and semistructured GID interview indicated that she was gender dysphoric at follow-up.

In the Deogracias et al. (2007) study, a cutoff score of  $\leq 3.00$  was used to indicate “caseness” for gender dysphoria on the GIDQ-AA. The 2 participants classified as gender dysphoric (and who completed the GIDQ-AA) had scores lower than 3.00 (means of 2.19 and 2.26, respectively), whereas the 18 participants classified as not gender dysphoric (and who completed the GIDQ-AA) all had scores 3.00 ( $M = 4.78$ ,  $SD = .20$ ; range, 4.30–5.00). There was a significant difference between these two subgroups,  $t(18) = 17.81$ ,  $p < .001$ ,  $d = 13.27$ , which supports the classification of the participants on the basis of the clinical interview.

Bakker et al. (1993) estimated that 1 in 30,400 genetically female adults in the general population have GID. Using this baseline prevalence value, the odds of persistent gender dysphoria (12%) in the present sample was 4,084 times the odds of gender dysphoria in the general population of biological females.

### Sexual Orientation

On the basis of the Kinsey interview ratings, participants were classified into the following three sexual orientation groups for fantasy and behavior: (a) heterosexual (Kinsey ratings of 0–1), (b) bisexual/homosexual (Kinsey ratings of 4–6), and (c) no sexual fantasy or behavior. For the fantasy ratings (see Table 3), 15

participants (60%) were classified as exclusively heterosexual, 8 (32%) were classified as bisexual/homosexual, and the remaining 2 (8%) were classified as having no sexual fantasies. Of the 3 participants classified as gender dysphoric, 2 were exclusively homosexual in fantasy (i.e., sexually attracted to members of their own birth sex). The other gender dysphoric participant reported no sexual fantasies and described herself as being “dead sexually.” (Of the 4 participants with a DSD, 3 were classified as exclusively heterosexual in fantasy, and 1 reported no sexual fantasies; 2 were classified as exclusively heterosexual in behavior, and 2 reported no sexual behavior.)

For the EROS, we compared the participants classified as exclusively heterosexual with those classified as bisexual/homosexual on the basis of their Kinsey ratings. With age at follow-up covaried, a 2 (sexual orientation: heterosexual vs. bisexual/homosexual)  $\times$  2 (EROS: attraction to men vs. attraction to women) analysis of covariance (ANCOVA) revealed a significant Sexual Orientation  $\times$  EROS interaction,  $F(1, 20) = 25.67, p < .001$ , partial  $\eta^2 = .56$ .

Independent  $t$  tests showed that participants classified as heterosexual in fantasy had, on average, a higher heteroerotic EROS score ( $M = 2.03, SD = .87$ ) than participants classified as bisexual/homosexual in fantasy ( $M = 1.84, SD = 1.34$ ), but the difference was not significant,  $t(20) < 1, d = .19$ ; however, participants classified as bisexual/homosexual reported, on average, a significantly higher EROS homoerotic score ( $M = 3.32, SD = 1.25$ ) than participants classified as heterosexual ( $M = 1.02, SD = .07$ ),  $t(20) = -7.28, p < .001, d = -3.33$ . A paired-samples  $t$  test was conducted to evaluate whether participants classified as heterosexual reported higher heteroerotic fantasies than homoerotic fantasies. The results indicated that the mean heteroerotic score was significantly greater than the mean homoerotic score,  $t(14) = 4.75, p < .001$ , with a large effect size of 1.23. Conversely, participants classified as bisexual/homosexual reported significantly higher homoerotic fantasies than heteroerotic fantasies,  $t(6) = -2.61, p < .04$ , with a large effect size of  $-0.99$ .

Regarding Kinsey ratings of sexual orientation in behavior (see Table 3), 11 participants (44%) were classified as exclusively heterosexual, 6 (24%) were classified as bisexual/homosexual, and the remaining 8 (32%) were classified as having no sexual experiences. Of the 3 participants classified as gender dysphoric, 2 were exclusively homosexual in behavior (i.e., had sexual experiences with members of their own birth sex). The other gender dysphoric participant reported no sexual behaviors.

For the SHQ ratings, we compared the participants classified as exclusively heterosexual with those classified as bisexual/homosexual on the basis of their Kinsey ratings. A 2 (sexual orientation: heterosexual vs. bisexual/homosexual)  $\times$  2 (SHQ: with men vs. with women) analysis of variance (ANOVA) revealed a significant Sexual Orientation  $\times$  SHQ interaction,  $F(1, 13) = 70.41, p < .001$ , partial  $\eta^2 = .84$ . Independent  $t$  tests for the SHQ scores showed that participants classified as heterosexual in behavior reported, on average, significantly more heterosexual sexual experiences ( $M = 2.15, SD = .54$ ) than participants classified as bisexual/homosexual ( $M = 1.00, SD = .00$ ),  $t(13) = 4.12, p = .001, d = 2.42$ . In fact, participants classified as bisexual/homosexual reported no sexual experiences with men over the past 12 months. Participants classified as bisexual/homosexual reported, on average, significantly more homosexual sexual experi-

ences ( $M = 2.48, SD = .40$ ) than did participants classified as heterosexual ( $M = 1.04, SD = .12$ ),  $t(13) = -11.17, p < .001, d = -6.56$ .

For participants classified as having a “typical” (i.e., non-gender-dysphoric) gender identity at follow-up, there were no substantive disjunctions between Kinsey ratings and sexual identity self-labeling (see Table 3). One exception was a participant who self-labeled as heterosexual, although she did not report any sexual fantasies or behaviors in the 12 months prior to the interview. For the 3 participants classified as gender dysphoric at follow-up, 2 self-labeled as heterosexual; however, it should be noted that their sexual orientation in relation to their birth sex was homosexual. As noted earlier, the remaining gender-dysphoric participant felt that she was “dead sexually” and labeled herself as asexual.

One participant (ID 5 in Table 3) was classified as bisexual/homosexual in fantasy but heterosexual in behavior. Her self-labeled sexual identity was bisexual. For the 17 participants who could be assigned a Kinsey rating between 0 and 6 for both behavior and fantasy (i.e., excluding the 8 individuals who did not report any sexual behavior [ $n = 6$ ] or any sexual fantasy and behavior [ $n = 2$ ]; see Table 3), the correlation between Kinsey fantasy and behavior ratings was  $.93 (df = 15), p < .001$ .

#### *Odds Ratios for Bisexual/Homosexual Sexual Orientation in Fantasy and Behavior*

Odds ratios were calculated for bisexual/homosexual sexual orientation in fantasy and behavior using prevalence estimates from several major survey studies of sexual orientation in adolescent girls and young women (Dickson, Paul, & Herbison, 2003; Fergusson, Horwood, Ridder, & Beautrais, 2005; McCabe, Hughes, Bostwick, & Boyd, 2005; Narring, Stronski, & Michaud, 2003; Remafedi, Resnick, Blum, & Harris, 1992; Russell & Seif, 2002). From these studies, base rates for bisexual/homosexual sexual orientation in fantasy and behavior were estimated to range from 2.0% to 5.0% in the female general population. The odds of reporting bisexual/homosexual sexual orientation in fantasy in the present sample was 8.9–23.1 times higher, and the odds of reporting bisexual/homosexual sexual orientation in behavior in the present sample was 6.0–15.5 times higher than it is in women in the general population.

#### *Relation Between Age and Sexual Orientation*

Table 4 shows the means and standard deviations of ages at assessment and at follow-up as a function of Kinsey groups in fantasy and behavior, respectively. For the Kinsey fantasy ratings, a one-way ANOVA for age at follow-up was significant,  $F(2, 22) = 4.91, p = .017$ , while the ANOVA for age at assessment in childhood approached statistical significance,  $F(2, 22) = 2.58, p = .098$ . At follow-up, participants classified as bisexual/homosexual were, on average, significantly older than participants classified as heterosexual or asexual,  $t(21) = -2.54, p = .019$ , and  $t(8) = -2.37, p = .046$ , respectively. There was no significant difference in the mean age at follow-up between participants classified as heterosexual and those classified as asexual,  $t(15) = -1.30, p = .211$ . For the Kinsey behavior ratings, the one-way ANOVAs for

Table 4  
*Means and Standard Deviations of Age (in Years) as a Function of Kinsey Ratings in Fantasy and Behavior*

Age	None		Exclusively heterosexual		Bisexual/homosexual		<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
By Kinsey fantasy ratings <sup>a</sup>							
At assessment	12.42	.35	7.96	3.10	9.75	2.73	.098
At follow-up	17.22	.18	21.76	4.78	27.50	5.88	.017
By Kinsey behavior ratings <sup>b</sup>							
At assessment	9.94	3.99	7.51	2.51	10.02	1.91	.142
At follow-up	20.66	5.11	22.81	4.79	27.45	6.93	.087

<sup>a</sup> For participants grouped by Kinsey fantasy ratings,  $n = 2$ ,  $n = 15$ , and  $n = 8$  for participants with no fantasies, exclusively heterosexual fantasies, and bisexual/homosexual fantasies, respectively. <sup>b</sup> For participants grouped by Kinsey behavior ratings,  $n = 8$ ,  $n = 11$ , and  $n = 6$  for participants with no behaviors, exclusively heterosexual behaviors, and bisexual/homosexual behaviors, respectively.

age at assessment and follow-up were nonsignificant,  $F(2, 22) = 2.14$ ,  $p = .142$ , and  $F(2, 22) = 2.73$ ,  $p = .087$ , respectively.

#### *Social Desirability*

One-way ANCOVAs (age at follow-up covaried) were conducted to evaluate the proportion of socially desirable responses on the M–C SDS for participants classified as heterosexual, bisexual/homosexual, and asexual in fantasy and behavior. There were no significant differences in the proportion of socially desirable responses on the M–C SDS as a function of Kinsey ratings in either fantasy or behavior,  $F(2, 20) = 1.00$ ,  $ns$ , and  $F(2, 20) < 1$ , respectively (data not shown; see footnote 2).

#### *Relation Between Sex-Typed Child Behavior and Sexual Orientation*

To evaluate whether degree of cross-sex-typed behavior in childhood was related to sexual orientation at follow-up, we used the  $z$ -composite of sex-typed behavior as a function of Kinsey classification in fantasy (heterosexual, bisexual/homosexual, asexual). With age at follow-up covaried, there was no significant difference in participants' cross-sex-typed behavior in childhood as a function of sexual orientation in fantasy,  $F(2, 21) = 1.06$ , partial  $\eta^2 = .09$  (data not shown; see footnote 2). For Kinsey ratings in behavior, however, a one-way ANCOVA was significant,  $F(2, 21) = 6.45$ ,  $p = .006$ , the strength of which was large, as assessed by partial  $\eta^2$ , with the Kinsey ratings accounting for 37% of the variance of participants' cross-sex-typed behavior in childhood. Participants classified as bisexual/homosexual ( $M = .52$ ,  $SD = .49$ ) had significantly more cross-sex-typed behavior in childhood than participants classified as heterosexual ( $M = -.04$ ,  $SD = .45$ ) or asexual ( $M = -.33$ ,  $SD = .39$ ), both  $ps < .05$ . There was no significant difference in the mean  $z$ -composite of sex-typed child behavior between participants classified as heterosexual and those classified as asexual (see footnote 2).

For the Kinsey ratings in behavior, we reran this analysis with the 3 gender-dysphoric participants removed (2 were classified as bisexual/homosexual and 1 was classified as asexual). For the

$z$ -composite, the main effect for Kinsey ratings in behavior remained statistically significant,  $F(2, 18) = 3.58$ ,  $p < .05$ , partial  $\eta^2 = .29$ .

#### *Relation Between Recalled Childhood Cross-Gender Behavior and Gender Identity at Follow-Up*

We conducted an evaluation of recalled cross-gender behavior between gender-dysphoric and non-gender-dysphoric participants. Table 5 shows the means and standard deviations of the RCGI Factor 1 score. Participants classified as gender dysphoric at follow-up ( $n = 2$ ;  $Ms = 1.29$  and  $1.81$ , respectively) recalled significantly more cross-gender identity and role behavior in childhood than participants classified as having no gender dysphoria

Table 5  
*Mean Factor Scores and Standard Deviations on the Recalled Childhood Gender Identity/Gender Role Questionnaire (Zucker et al., 2006) for Gender Identity Status and Sexual Orientation at Follow-Up*

Group	<i>M</i>	<i>SD</i>	<i>d</i>	<i>n</i>
Gender identity status				
Gender dysphoric	1.55	.36	-1.96	2
No gender dysphoria (Adolescent girls with GID)	2.69 (2.15)	.59 (.58)		18 (25)
Sexual orientation <sup>a</sup>				
Heterosexual (Heterosexual comparison sample)	2.82 (3.34)	.54 (.53)	1.88	15 (30)
Bisexual/homosexual (Homosexual comparison sample)	1.84 (2.68)	.44 (.72)		5 (21)

*Note.* The absolute range was 1.00–5.00. A lower score indicates more recalled atypical gender identity and gender role behavior. Twenty participants completed the questionnaire because the RCGI was not yet part of the follow-up protocol for 5 participants. Groups and values in parentheses are from Zucker et al. (2006); the factor scores were from a sample of heterosexual and homosexual female university students unselected for gender identity. GID = gender identity disorder.

<sup>a</sup> Sexual orientation was determined on the basis of Kinsey ratings for fantasy and behavior.



( $n = 18$ ;  $M = 2.69$ ; range, 1.56–3.87),  $t(18) = -2.62$ ,  $p = .017$ . As shown in Table 5, the mean Factor 1 score on the RCGI for the participants with persistent gender dysphoria was more extreme than it was for a sample of clinic-referred adolescent girls ( $n = 25$ ) with GID reported on by Zucker et al. (2006), whereas the mean score of the participants without gender dysphoria was somewhat less extreme.

Further analyses on the RCGI Factor 1 score by sexual orientation revealed that participants classified as bisexual/homosexual recalled significantly more cross-gender identity and gender role behavior in childhood than did participants classified as heterosexual or asexual,  $t(18) = 3.65$ ,  $p = .002$ .

## Discussion

The data reported in this article represent the first systematic psychosexual follow-up into late adolescence and young adulthood of clinic-referred girls with potential problems in their gender identity development. The two key findings were as follows: (a) the percentage of girls with persistent gender dysphoria was modest but arguably higher than the base rate of GID in the general population of biological females, and (b) the percentage of girls who differentiated a later bisexual/homosexual sexual orientation was moderate but clearly higher than the base rates of bisexual/homosexual sexual orientation in general survey and epidemiological studies of adolescent girls and young adult women in which sexual orientation (in fantasy and/or behavior) was assessed with at least some gradation in response options (as opposed to simple dichotomous items).

Before providing an analysis of these findings, we note two limitations of the study. First, the sample size was small, but this is, at least in part, understandable because the number of referred girls to specialized gender identity clinics is notably lower than that of referred boys (e.g., Cohen-Kettenis et al., 2003, 2006). Second, the present study did not have a concurrent control group (e.g., a group of girls referred for other kinds of clinical concerns or a group of nonreferred girls). Accordingly, some of our comparative analyses relied on epidemiological or survey data.

Regarding the persistence of gender dysphoria from the childhood assessment to the follow-up, the present study found that the vast majority of the girls showed desistance: 88% of the girls did not report distress about their gender identity at follow-up. The high rate of desistance appears to differ quite markedly from the findings of other follow-up studies of adolescent girls and adult women with GID (in which the baseline assessment is in adolescence or adulthood). In these studies, the rate of GID persistence appears to be, at minimum, around 70% (Cohen-Kettenis & van Goozen, 1997; Smith, van Goozen, Kuiper, & Cohen-Kettenis, 2005). In a comparative developmental perspective, then, there appears to be important variation in GID persistence between childhood and adolescence/young adulthood.

How might this disjunction be understood? One possibility pertains to the differences in the *DSM* criteria for GID that are used for children versus those that are used for adolescents/adults. The criteria for GID in girlhood place relatively greater weight on surface behaviors of cross-gender identification, whereas the criteria in adolescence and adulthood rely more strongly on behaviors and feelings pertaining to the disjunction between gender subjectivity and somatic sex. Thus, it is conceivable that the childhood

criteria for GID may “scoop in” girls who are at relatively low risk for adolescent/adult gender dysphoria, which revolves so much around somatic indicators (e.g., distress regarding breast development or other markers of physical femaleness, etc.).

It should, however, be noted that adolescent girls and adult women with GID typically recall the same kinds of cross-gender behavior patterns in girlhood that correspond to the *DSM* criteria for GID in childhood (e.g., Blanchard & Freund, 1983; Pearlman, 2006; Zucker et al., 2006), which are then augmented and exacerbated by the external physical markers of biological femaleness at puberty. Indeed, in the present study, the recalled sex-typed behavior from childhood of our participants was reasonably similar to the childhood recollections of girls with GID assessed for the first time in adolescence (see Table 5).

In the present study, 40% of the girls were not judged to have met the complete *DSM* criteria for GID at the time of childhood assessment (although some of these girls likely had met the complete criteria at some earlier point in their development). Thus, on the one hand, it could be argued that if some of the girls were subthreshold for GID in childhood, then one might assume that they would not be at risk for GID in adolescence or adulthood. On the other hand, it could be argued that cross-gender identification in girlhood (including subthreshold GID) is a risk factor for later GID; that is, under some conditions, there is an intensification of cross-gender identification that results in the development of gender dysphoria (see Green, 2003). Indeed, clinical experience with adolescent girls with GID indicates that not all of them would have met the complete criteria for GID in girlhood. Indeed, it is not uncommon for the parents of these girls to recall that their daughters identified as “tomboys” during childhood and that they did not remember them voicing the desire to want to become a boy, but that their gender dysphoria emerged only around the time of puberty (see, e.g., Pearlman, 2006; Zucker, 2006, Case 1).

If one accepts the argument that girlhood cross-gender identification is a risk factor for gender dysphoria in adolescence and adulthood, the relatively high rate of desistance in the current study (in comparison with the relatively high rate of persistence seen in gender-dysphoric girls and women assessed for the first time in adolescence or adulthood) suggests that there is some type of plasticity in gender identity differentiation that operates early in development but then narrows considerably by adolescence. Thus, at least among the girls in the present sample, some factor or set of factors may have operated to lessen the likelihood that their gender dysphoria or cross-gender identification would persist or intensify in adolescence and adulthood. Of course, such factors could include both biological and psychosocial influences, but the systematic identification of such factors was beyond the scope of the present investigation.

To our knowledge, the results of the present study represent the first prospective data set that shows that girlhood cross-gender identification is associated with a relatively high rate of bisexual/homosexual sexual orientation in adolescence and adulthood. Using survey data on sexual orientation in young women as a comparative metric, we estimated that the odds of reporting a bisexual/homosexual sexual orientation in fantasy was 8.9–23.1 times higher in the present sample and that the odds of reporting a bisexual/homosexual sexual orientation in behavior was 6.7–15.5 times higher. In this respect, the data show at least some conver-

gence with data from retrospective studies (Bailey & Zucker, 1995).

A strength of the present study was that the assessment of sexual orientation was based on a multiparameter, face-to-face interview from which Kinsey global ratings were derived and that was complemented by self-report on psychometrically sound questionnaires (cf. Savin-Williams, 2006). Although one has to be cautious about the possibility that our participants underreported a minority sexual orientation, it should be recalled that we found no significant relation between our Kinsey classifications and the propensity to give socially desirable responses on the M-C SDS.

Our classification of participants' sexual orientation was based on fantasy and behavior ratings for the 12-month period prior to follow-up. In the literature on women's sexual orientation, there has been a lot of recent discussion regarding its stability versus its fluidity (see, e.g., Baumeister, 2000; Peplau et al., 1999). Diamond (2005b), for example, followed 79 self-labeled lesbian, bisexual, and "unlabeled" sexual minority women over an 8-year period (mean age at baseline, 19 years). At the 8-year follow-up, 92.4% of the women continued to self-label as lesbian, bisexual, or unlabeled, although there was considerable fluctuation within these three categories over time (e.g., lesbian to bisexual or unlabeled to lesbian). The remaining 7.5% of the women self-labeled as heterosexual at the follow-up. In our view, Diamond's (2005b) data suggest considerable stability of a minority sexual orientation despite the evidence of greater fluidity within the subcategories of lesbian, bisexual, and unlabeled.

One limitation of Diamond's study was that it did not include a group of self-labeled heterosexual women at baseline; thus, comparative evidence on the stability or fluidity of a majority sexual orientation was not available. Using data from the National Longitudinal Survey of Adolescent Health, however, Savin-Williams and Ream (2007) provided data on the stability of a heterosexual sexual orientation (attraction and behavior) of several thousand girls and women between the ages of 15 and 26 years in a three-wave assessment. In their study, there was considerable evidence for a stable heterosexual sexual orientation. For example, only 3.1% of girls who reported exclusive heterosexual attractions at Wave 1 reported bisexual or lesbian attractions at Wave 3, and only 3.5% of girls who reported exclusive heterosexual behavior at Wave 1 reported bisexual or lesbian behavior at the Wave 3. Given these findings, the case could be made that our participants' sexual orientations will remain relatively stable over time but, on this point, only continued follow-up can test this conjecture empirically.

Because there was considerable variability in sexual orientation at follow-up, we made some relatively crude efforts at predicting such variation (compromised, of course, by the small sample size). There were hints in the data that younger age at assessment in childhood was associated with a later heterosexual sexual orientation (Table 4), but the effects were weak. The composite index of sex-typed behavior in childhood was not significantly associated with sexual orientation in fantasy, but it was with sexual orientation in behavior, with those participants classified as bisexual/homosexual exhibiting more cross-gender behavior. We also found that participants classified as bisexual/homosexual recalled having engaged in more cross-gender behavior during childhood than those classified as heterosexual or asexual (Table 5). These data are suggestive, therefore, of a "dosage" effect, that is, that

degree of girlhood cross-gender identification is associated with a greater likelihood of a later minority sexual orientation. Of course, these preliminary findings need to be confirmed in much larger clinical samples; in addition, it would be desirable to examine whether or not variation in degree of girlhood cross-sex-typed behavior is related to sexual orientation in epidemiological samples drawn from nonclinical populations.

How do the results of the present study compare with those of follow-up studies of boys with GID? In Zucker (2005b), a follow-up on 40 boys with GID from the same clinic, using the same methods as in the present study, showed a persistence rate of 20%, only modestly higher than the rate of 12% for the girls in the present study. In Zucker (2005b), 42.5% of the boys were classified as bisexual/homosexual in fantasy, which is again only modestly higher than the rate of 32% for the girls in the present study; however, the rate of a bisexual/homosexual sexual orientation in fantasy was considerably lower than the 75% found by Green (1987) in his study of feminine boys. In comparison with the boys followed up by Green (1987) and by Zucker (2005b), it is important to note that the girls in the present study were, on average, several years older at follow-up, which, if anything, would suggest that the likelihood of underreporting a minority sexual orientation would be lower for this sample than for the samples of boys.

If it proves to be the case that cross-sex-typed behavior is, indeed, less closely linked to a later bisexual/homosexual sexual orientation in girls than it is in boys, this would be consistent with a prediction made by Bailey and Zucker (1995) in their meta analytic retrospective study. It would also be consistent with recent theorizing on the greater flexibility of sexual orientation in women, in which it has been argued that relational factors during adolescence and adulthood play a more important role in sexual partner preference than it does in men (Diamond, 2003, 2005a; Peplau et al., 1999). It is apparent from the present study that there is considerable within-sex variation to be explained in the long-term psychosexual differentiation of behaviorally masculine girls, with regard to both gender identity and sexual orientation. These findings suggest that any reductionist account of psychosexual differentiation will likely be unable to capture this variation. Multivariate models are clearly required in order to identify the best predictors of such within-sex variation. On this point, the field will hopefully move forward as larger samples are collated, including prospective, epidemiologically based cohorts that incorporate theoretically based predictor variables.

## References

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders* (3rd ed.). Washington, DC: Author.
- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed., rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.
- Bailey, J. M., Bechtold, K. T., & Berenbaum, S. A. (2002). Who are tomboys and why should we study them? *Archives of Sexual Behavior, 31*, 333-341.
- Bailey, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology, 31*, 43-55.

- Bakker, A., van Kesteren, P. J. M., Gooren, L. J. G., & Bezemer, P. D. (1993). The prevalence of transsexualism in the Netherlands. *Acta Psychiatrica Scandinavica*, *87*, 237–238.
- Barrett, A. E., & White, H. R. (2002). Trajectories of gender role orientation in adolescence and early adulthood: A prospective study of the mental health effects of masculinity and femininity. *Journal of Health and Social Behavior*, *43*, 451–468.
- Bartlett, N. H., & Vasey, P. L. (2006). A retrospective study of childhood gender-atypical behavior in Samoan *Fa'afafine*. *Archives of Sexual Behavior*, *35*, 659–666.
- Bates, J. E., & Bentler, P. M. (1973). Play activities of normal and effeminate boys. *Developmental Psychology*, *9*, 20–27.
- Baumeister, R. F. (2000). Gender differences in erotic plasticity: The female sex drive as socially flexible and responsive. *Psychological Bulletin*, *126*, 347–374.
- Berenbaum, S. A., & Bailey, J. M. (2003). Effects on gender identity of prenatal androgens and genital appearance: Evidence from girls with congenital adrenal hyperplasia. *Journal of Clinical Endocrinology & Metabolism: Clinical and Experimental*, *88*, 1101–1106.
- Blanchard, R., & Freund, K. (1983). Measuring masculine gender identity in females. *Journal of Consulting and Clinical Psychology*, *51*, 205–214.
- Cohen-Kettenis, P. T., Owen, A., Kaijser, V. G., Bradley, S. J., & Zucker, K. J. (2003). Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: A cross-national, cross-clinic comparative analysis. *Journal of Abnormal Child Psychology*, *31*, 41–53.
- Cohen-Kettenis, P. T., & Pfäfflin, F. (2003). *Transgenderism and intersexuality in childhood and adolescence: Making choices*. Thousand Oaks, CA: Sage.
- Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1997). Sex reassignment of adolescent transsexuals: A follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 263–271.
- Cohen-Kettenis, P. T., Wallien, M., Johnson, L. L., Owen-Anderson, A. F. H., Bradley, S. J., & Zucker, K. J. (2006). A parent-report Gender Identity Questionnaire for Children: A cross-national, cross-clinic comparative analysis. *Clinical Child Psychology and Psychiatry*, *11*, 397–405.
- Cohler, B. J., & Galatzer-Levy, R. M. (2000). *The course of gay and lesbian lives: Social and psychoanalytic perspectives*. Chicago: University of Chicago Press.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*, 349–354.
- Deogracias, J. J., Johnson, L. L., Meyer-Bahlburg, H. F. L., Kessler, S. J., Schober, J. M., & Zucker, K. J. (2007). The Gender Identity/Gender Dysphoria Questionnaire for Adolescents and Adults. *Journal of Sex Research*, *44*, 370–379.
- Diamond, L. M. (2003). What does sexual orientation orient? A biobehavioral model distinguishing romantic love and sexual desire. *Psychological Review*, *110*, 173–192.
- Diamond, L. M. (2005a). From the heart or the gut? Sexual-minority women's experiences of desire for same-sex and other-sex partners. *Feminism & Psychology*, *15*, 10–14.
- Diamond, L. M. (2005b). A new view of lesbian subtypes: Stable versus fluid identity trajectories over an 8-year period. *Psychology of Women Quarterly*, *29*, 119–128.
- Dickson, N., Paul, C., & Herbison, P. (2003). Same-sex attraction in a birth cohort: Prevalence and persistence in early adulthood. *Social Science & Medicine*, *56*, 1607–1615.
- Doorn, C. D., Poortinga, J., & Verschoor, A. M. (1994). Cross-gender identity in transvestites and male transsexuals. *Archives of Sexual Behavior*, *23*, 185–201.
- Drummond, K. D. (2006). *A follow-up study of girls with gender identity disorder*. Unpublished master's thesis, Ontario Institute for Studies in Education of the University of Toronto, Toronto, Ontario, Canada.
- Ehrhardt, A. A., Grisanti, G., & McCauley, E. A. (1979). Female-to-male transsexuals compared to lesbians: Behavioral patterns of childhood and adolescent development. *Archives of Sexual Behavior*, *8*, 481–490.
- Fergusson, D. M., Horwood, L. J., Ridder, E. M., & Beautrais, A. L. (2005). Sexual orientation and mental health in a birth cohort of young adults. *Psychological Medicine*, *35*, 971–981.
- Freund, K., Langevin, R., Satterberg, J., & Steiner, B. (1977). Extension of the Gender Identity Scale for Males. *Archives of Sexual Behavior*, *6*, 507–519.
- Fridell, S. R., Owen-Anderson, A., Johnson, L. L., Bradley, S. J., & Zucker, K. J. (2006). The Playmate and Play Style Preferences Structured Interview: A comparison of children with gender identity disorder and controls. *Archives of Sexual Behavior*, *35*, 729–737.
- Gottschalk, L. (2003). Same-sex sexuality and childhood gender non-conformity: A spurious connection. *Journal of Gender Studies*, *12*, 35–50.
- Green, R. (1987). *The "sissy boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.
- Green, R. (2003). The "T" word [Letter to the editor]. *Archives of Sexual Behavior*, *32*, 1.
- Green, R., Williams, K., & Goodman, M. (1982). Ninety-nine "tomboys" and "non-tomboys": Behavioral contrasts and demographic similarities. *Archives of Sexual Behavior*, *11*, 247–266.
- Hardt, J., & Rutter, M. (2004). Validity of adult retrospective reports of adverse childhood experiences: Review of the evidence. *Journal of Child Psychology and Psychiatry*, *45*, 260–273.
- Hegarty, P. (1999). *Recalling childhood, constructing identity: Norms and stereotypes in theories of sexual orientation*. Unpublished doctoral dissertation, Stanford University, Stanford, CA.
- Holden, R. R., & Fekken, G. C. (1989). Three common social desirability scales: Friends, acquaintances, or strangers? *Journal of Research in Personality*, *23*, 180–191.
- Hollingshead, A. B. (1975). *Four factor index of social status*. Unpublished manuscript, Department of Sociology, Yale University, New Haven, CT.
- Hughes, I. A., Houk, C., Ahmed, S. F., Lee, P. A., & Lawson Wilkins Pediatric Endocrine Society/European Society for Paediatric Endocrinology Consensus Group. (2006). Consensus statement on management of intersex disorders. *Archives of Disease in Childhood*, *91*, 554–563.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*, 581–592.
- Johnson, L. L., Bradley, S. J., Birkenfeld-Adams, A. S., Radzins Kuksis, M. A., Maing, D. M., Mitchell, J. N., & Zucker, K. J. (2004). A parent-report Gender Identity Questionnaire for Children. *Archives of Sexual Behavior*, *33*, 105–116.
- Khuri, J., & Ruble, D. N. (2006, April). *The consequences of early gender-typicality: Findings in a female sample*. Poster session presented at the Second Gender Development Conference, San Francisco, CA.
- King, M. F., & Brunner, G. C. (2000). Social desirability bias: A neglected aspect of validity testing. *Psychology and Marketing*, *17*, 79–103.
- Kinsey, A. C., Pomeroy, W. B., & Martin, C. E. (1948). *Sexual behavior in the human male*. Philadelphia: W. B. Saunders.
- Kite, M. E., & Deaux, K. (1987). Gender belief systems: Homosexuality and the implicit inversion theory. *Psychology of Women Quarterly*, *11*, 83–96.
- Langevin, R. (1985). *Sexual strands: Understanding and treating sexual anomalies in men*. Hillsdale, NJ: Erlbaum.
- Laumann, E. O., Gagnon, J. H., Michael, R. T., & Michaels, S. (1994). *The social organization of sexuality: Sexual practices in the United States*. Chicago: University of Chicago Press.
- Lippa, R. (1998). Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology*, *74*, 996–1009.



- Lippa, R. A. (2002). *Gender, nature, and nurture*. Mahwah, NJ: Erlbaum.
- Liss, M. B. (Ed.). (1983). *Social and cognitive skills: Sex roles and children's play*. New York: Academic Press.
- Maccoby, E. E. (1998). *The two sexes: Growing up apart, coming together*. Cambridge, MA: Belknap Press.
- McCabe, S. E., Hughes, T. L., Bostwick, W., & Boyd, C. J. (2005). Assessment of difference in dimensions of sexual orientation: Implications for substance use research in a college-age population. *Journal of Studies on Alcohol, 66*, 620–629.
- Meyer-Bahlburg, H. F. L. (1994). Intersexuality and the diagnosis of gender identity disorder. *Archives of Sexual Behavior, 23*, 21–40.
- Meyer-Bahlburg, H. F. L. (2005). Gender identity outcome in female-raised 46,XY persons with penile agenesis, cloacal exstrophy of the bladder, or penile ablation. *Archives of Sexual Behavior, 34*, 423–438.
- Narring, F., Stronski, H. S., & Michaud, P.-A. (2003). Prevalence and dimensions of sexual orientation in Swiss adolescents: A cross-sectional survey of 16- to 20-year-old students. *Acta Paediatrica, 92*, 233–239.
- Pearlman, S. F. (2006). Terms of connection: Mother-talk about female-to-male transgender children. *Journal of GLBT Family Studies, 2*(3/4), 93–122.
- Peplau, L. A., & Huppini, M. (in press). Masculinity, femininity and the development of sexual orientation. *Journal of Gay and Lesbian Psychotherapy*.
- Peplau, L. A., Spalding, L. R., Conley, T. D., & Veniegas, R. C. (1999). The development of sexual orientation in women. *Annual Review of Sex Research, 10*, 70–99.
- Remafedi, G., Resnick, M., Blum, R., & Harris, L. (1992). Demography of sexual orientation in adolescents. *Pediatrics, 89*, 714–721.
- Ruble, D. N., Martin, C. L., & Berenbaum, S. A. (2006). Gender development. In W. Damon & R. M. Lerner (Series Eds.) and N. Eisenberg (Vol. Ed.), *Handbook of child psychology* (6th ed.), Vol. 3: *Social, emotional, and personality development* (pp. 858–932). New York: Wiley.
- Russell, S. T., & Seif, H. (2002). Bisexual female adolescents: A critical analysis of past research, and results from a national survey. *Journal of Bisexuality, 2*(2–3), 73–94.
- Savin-Williams, R. C. (2006). Who's gay? Does it matter? *Current Directions in Psychological Science, 15*, 40–44.
- Savin-Williams, R. C., & Ream, G. L. (2007). Prevalence and stability of sexual orientation components during adolescence and young adulthood. *Archives of Sexual Behavior, 36*, 385–393.
- Silverthorn, N. A., & Gekoski, W. L. (1995). Social desirability effects on measures of adjustment to university, independence from parents, and self-efficacy. *Journal of Clinical Psychology, 51*, 244–251.
- Smith, Y. L. S., van Goozen, S. H. M., Kuiper, A. J., & Cohen-Kettenis, P. T. (2005). Sex reassignment: Outcomes and predictors of treatment for adolescent and adult transsexuals. *Psychological Medicine, 35*, 89–99.
- Storms, M. D. (1980). Theories of sexual orientation. *Journal of Personality and Social Psychology, 38*, 783–792.
- Strahan, R., & Gerbasi, K. C. (1972). Short, homogeneous versions of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 28*, 191–193.
- Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.
- Wechsler, D. (1974). *Manual: Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corporation.
- Wechsler, D. (1991). *Wechsler Intelligence Scale for Children—Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (1997). *Wechsler Adult Intelligence Scale—Third Edition*. San Antonio, TX: Psychological Corporation.
- Wechsler, D. (2002). *Wechsler Preschool and Primary Scale of Intelligence—Third Edition*. San Antonio, TX: Psychological Corporation.
- Zucker, K. J. (2005a). Commentary on Gottschalk's (2003) "Same-sex sexuality and childhood gender non-conformity: A spurious connection." *Journal of Gender Studies, 14*, 55–60.
- Zucker, K. J. (2005b). Gender identity disorder in children and adolescents. *Annual Review of Clinical Psychology, 1*, 467–492.
- Zucker, K. J. (2005c). Measurement of psychosexual differentiation. *Archives of Sexual Behavior, 34*, 375–388.
- Zucker, K. J. (2006). Gender identity disorder. In D. A. Wolfe & E. J. Mash (Eds.), *Behavioral and emotional disorders in adolescents: Nature, assessment, and treatment* (pp. 535–562). New York: Guilford Press.
- Zucker, K. J. (in press). Reflections on the relation between sex-typed behavior in childhood and sexual orientation in adulthood. *Journal of Gay & Lesbian Psychotherapy*.
- Zucker, K. J., & Bradley, S. J. (1995). *Gender identity disorder and psychosexual problems in children and adolescents*. New York: Guilford Press.
- Zucker, K. J., Bradley, S. J., Lowry Sullivan, C. B., Kuksis, M., Birkenfeld-Adams, A., & Mitchell, J. N. (1993). A gender identity interview for children. *Journal of Personality Assessment, 61*, 443–456.
- Zucker, K. J., Bradley, S. J., Oliver, G., Blake, J., Fleming, S., & Hood, J. (1996). Psychosexual development of women with congenital adrenal hyperplasia. *Hormones and Behavior, 30*, 300–318.
- Zucker, K. J., Doering, R. W., Bradley, S. J., & Finegan, J. K. (1982). Sex-typed play in gender-disturbed children: A comparison to sibling and psychiatric controls. *Archives of Sexual Behavior, 11*, 309–321.
- Zucker, K. J., Finegan, J. K., Doering, R. W., & Bradley, S. J. (1983). Human figure drawings of gender-problem children: A comparison to siblings, psychiatric, and normal controls. *Journal of Abnormal Child Psychology, 11*, 287–298.
- Zucker, K. J., Lozinski, J. A., Bradley, S. J., & Doering, R. W. (1992). Sex-typed responses in the Rorschach protocols of children with gender identity disorder. *Journal of Personality Assessment, 58*, 295–310.
- Zucker, K. J., Mitchell, J. N., Bradley, S. J., Tkachuk, J., Cantor, J. M., & Allin, S. M. (2006). The Recalled Childhood Gender Identity/Gender Role questionnaire: Psychometric properties. *Sex Roles, 54*, 469–483.

Received September 27, 2006

Revision received July 24, 2007

Accepted July 31, 2007 ■