Boys will be Boys: The Effect of Social Evaluation Concerns on Gender-Typing

Robin Banerjee and Vicki Lintern, *University of Sussex, UK*

Abstract

Previous research has demonstrated that young children hold strong gender stereotypes for activities and toy preferences. Some researchers have argued that this rigid gender-typing displayed by young children is associated with peer reinforcement for stereotypical behaviour and punishment of counterstereotypical behaviour. The present study tests the hypothesis that the gender-typing displayed by young children is at least in part an active self-presentational effort to win positive evaluation from peers. Sixty-four children aged between 4 and 9 years described themselves in terms of their activity and toy preferences, once when alone and once when in front of a group of same-sex peers. They also completed a task measuring the rigidity of their gender stereotypes. It was found using both group-based and individual-based analyses that the children with the most rigid stereotypes—young boys—were more likely to present themselves as sex-typed in front of the peer audience than when alone. The older boys and the girls in all age groups tended to have less rigid stereotypes and their self-descriptions were in general not influenced by the presence of an audience. These results show that self-presentational concerns do influence children's gender-typed behaviour, and that these concerns may vary with age and gender.

Keywords: gender stereotypes; social evaluation; self-presentation; social cognition

Introduction

Children aged around 4 to 5 years have been shown to possess very strong gender stereotypes (e.g., Martin, 1989; Urberg, 1982). For example, Martin (1989) found that children of this age are unable to take into account counterstereotypical information when making judgements about people's toy preferences. Thus, even if a girl is said to like playing with cars, 4- to 6-year-olds will still insist that she would prefer to play with dolls than with aeroplanes. In contrast, older children are able to make judgements that take into account the counterstereotypical information they have been given, reflecting increases in the ability to coordinate the processing of categorical information (e.g., gender) and individuating information (e.g., personal toy preferences). This pattern of results is compatible with findings of a general decrease with age in gender bias (Powlishta, Serbin, Doyle, & White, 1994).

Much research has been directed towards determining the origins of this strong gender-typing in young children. One major focus has been on cognitive limitations.

Address for correspondence: Dr Robin Banerjee, School of Cognitive and Computing Sciences, University of Sussex, Falmer, Brighton, BN1 9QH, UK, Email: robinb@cogs.susx.ac.uk

For example, young children may have overly simple schematic representations of masculine and feminine characteristics that result in rigid adherence to gender stereotypes. Building on the foundation of theoretical work on gender schemas by Martin and Halverson (1981), Martin (1989) argues that young children's tendency to make stereotype-consistent inferences about behaviour may be the result of a heavy reliance on the categorical distinction between boys and girls. The failure of young children to take into account counterstereotypical individuating information when predicting another person's behaviour is consistent with this analysis: these children are not able to consider the implications of the counterstereotypical information and override the salient fact of the person's gender. Although this tendency to focus on the person's gender at the expense of individuating information can be seen broadly as a way of minimising information-processing demands, Bigler and Liben (1992) suggest that the specific cognitive limitation of the younger children here is an inability to classify target objects/people along multiple dimensions. Such multiple classification difficulties would lead to a failure to recognise simultaneously membership of a gender category and possession of a counterstereotypical attribute. The consequence of this is that young children demonstrate a rigid adherence to gender stereotypes in their judgements.

Such cognitive factors are coupled with social pressures that encourage stereotypical behaviour in young children. Many studies have shown that parents' gender stereotypes shape their treatment of boys and girls, right from birth. In these studies, parents are often found to describe newborn boys differently from newborn girls (Rubin, Provenzano, & Luria, 1974), and to encourage different activity and toy preferences. In the 'Baby X' studies, for example, male and female babies labelled as being of the other sex were treated by adults according to the gender stereotype suggested by the misleading label (e.g., Seavey, Katz, & Zalk, 1975).

Other social agents are also at work in gender development. Indeed, many parents explicitly try to avoid imposing gender stereotypes, but still fail to stop their children from acquiring and rigidly adhering to these stereotypes. One social agent of potentially great importance here is the child's peer group. Evidence has suggested that peers provide strong reinforcement for stereotypical behaviour while quickly showing their disapproval of counterstereotypical behaviour. Fagot (1977) for example observed that preschoolers respond negatively to peers who violate gender stereotypes. This appears to particularly true for boys, for whom being labelled a 'sissy' may be the ultimate insult. Moreover, this reinforcement of gender stereotypes is effective: children who are rewarded for stereotypical behaviour persist longer in that behaviour while children who are punished for counterstereotypical behaviour will tend to stop performing that behaviour (Lamb, Easterbrooks, & Holden, 1980; Lamb & Roopnarine, 1980). This is particularly true of the positive and negative contingencies established by same-sex peers (Fagot, 1985). The significance of peer presence on gender-typed play has been directly tested by Serbin, Connor, Burchardt, and Citron (1979), who found that children played less with counterstereotypical toys when in the presence of peers than when alone.

Little is known about the precise processes at work when we talk of these social pressures, and in particular about the active role of the child in responding to these pressures. The present research seeks to address this issue. Specifically, do children understand that they can *control* their gender-related behaviour in order to shape the way others see them? The present research compares children's gender-related self-descriptions when alone and when in the presence of a same-sex peer group. A dis-

tinction between the private and public self-descriptions would suggest that the child has an active role in responding to social pressures, and is tailoring his or her self-description in order to obtain positive social evaluation. Such discrimination between situations illustrates the important but often neglected point that children's gender-related behaviour may be, to some extent, a controlled public performance rather than an uncontrolled consequence of socialization experiences. Thus, children's rigid adherence to gender stereotypes may be regarded as a *social* phenomenon whose impact on behaviour is mediated by social motivations.

Research on children's worries suggests that social evaluation is not a general source of concern until around age 8 (Vasey, Crnic, & Carter, 1994), and indeed children are found to understand how social evaluation concerns can affect your behaviour from around age 8 (Banerjee & Yuill, 1999a). However, Banerjee & Yuill (1999b) have demonstrated that children as young as age 5 are cognitively capable of understanding self-presentational motives (which underlie behaviour that is designed to shape social evaluation). Thus, if a given domain is salient enough, children of this age may demonstrate concerns about social evaluation with regard to behaviour in that domain. Given that gender is an important part of children's self-concepts from a young age (see Golombok & Fivush, 1994), we might expect gender-related behaviour to be a domain where young children are concerned about social evaluation.

As we have seen, children at around the age of 5 have very rigid gender stereotypes, based on simplistic schemas about appropriate behaviours for boys and girls. Older children, on the other hand, are better able to move away from these gender stereotypes and take into account counterstereotypical preferences. Rigid adherence to gender stereotypes is likely to reflect a high salience of gender-related behaviours, and could therefore be related to social evaluation concerns. Following on from our argument above, we test in the present study the hypothesis that young children will be more likely than older children to alter their public self-descriptions to obtain positive social evaluation from their peers. First, we measure the extent of gender-typing in children's self-descriptions both in private and when in the presence of a same-sex peer group ('self-description task'). Also, we assess the rigidity of children's adherence to gender stereotypes when making judgements about boys and girls described as having counterstereotypical preferences ('judgement task'). Thus, the present study is able to test directly the link between children's cognitions about gender-related characteristics and their own self-presentational behaviour in social situations. Specifically, children who are very inflexible in their adherence to gender stereotypes when making judgements are expected to be most concerned about appearing sex-typed themselves in a public situation.

Method

Participants

The final sample consisted of 21 4- to 5-year-olds (10 girls, mean age 5;4, range 4;7–5;10), 21 6- to 7-year-olds (11 girls, *M* 6;5, range 6;1–7;5), and 22 8- to 9-year-olds (12 girls, *M* 9;1, range 8;6–9;6). Ten other children had failed to complete all three experimental sessions and were therefore excluded. The children were recruited from an urban primary school in a primarily working-class and ethnically diverse neighbourhood.

Measures

Two tasks were used in this experiment: a judgement task and a self-description task. Before describing these tasks, it should be noted that the selection of sex-typed toys and activities used in the tasks was informed by previous work on children's toy/activity preferences (e.g., De Lucia, 1963; Golombok & Rust, 1993). To provide further support for our classifications, twenty adults on a university campus were asked to rate how much boys and girls like to play with each item on a five-point scale (lower scores indicating more masculine-typed). All feminine-typed items had a mean of 4 or above, all masculine-typed items had a mean of 2 or below, and neutral items had a mean of 3.

Judgement task. This measure of rigidity in adherence to gender stereotypes when making judgements was derived from Martin (1989). It assesses children's ability to take into account counterstereotypical information about story characters in order to predict their toy preferences. As shown in Appendix A, each child heard about two male and two female target characters, in one of eight orders where the sex of the target character alternated. Two target characters had stereotypical interests (a boy whose best friend is a boy and who likes playing with aeroplanes, and a girl whose best friend is a girl and who likes playing with kitchen sets), while the other two had counterstereotypical interests (a boy whose best friend is a girl and who likes playing with toy prams/buggies, and a girl whose best friend is a boy and who likes playing with cars). This information was accompanied by a cartoon-style illustration of the target character with his or her preferred toy. For each target character, children were required to rate whether the character would like to play with dolls, play football, do skipping, and play with toy guns (all accompanied by cartoon-style drawings). These four items were fixed in an order which alternated from feminine to masculine. Ratings were made on a response sheet by circling a cross (to indicate NO), a tick (to indicate SOMETIMES), or two ticks (to indicate A LOT). The experimenters explained the rating system to the children, and initially showed them how to mark their answers. Children were soon comfortable with this rating system and were able to mark their own answers.

Self-description task. For this task, children were simply asked whether they played with particular toys or engaged in particular activities (each question accompanied by a cartoon-style illustration of the toy or activity). Five items were masculine-typed, five items were feminine-typed, and two were gender-neutral. Items were fixed in a random order, with the exception of the two neutral items which were placed at the beginning and half-way through the scale. Children used the same three-point rating scale as for the judgement task. Again, on the few occasions where difficulty with the rating system was encountered, children were given guidance. Children completed this task under two conditions. In the 'alone' condition, children completed the task alone with no other children present, and were simply told that they were to answer questions about what they enjoyed playing. In the 'group' condition, children completed the task sitting around a table in a group of 3 to 5 same-sex peers. Children were reminded that they had already completed the questionnaire the day before and were doing this for the second time, but were instructed that this time they were going to mark down their answers and show them to the other children at the table. For each item, all children in the group were required to mark an answer down and were then

asked to reveal their answer. The experimenter checked that each child marked an answer for each question. Cronbach's alpha was calculated for both masculine and feminine items, in each condition (alone and group). All alphas were above .74, and similarly high levels of internal consistency were obtained when the calculations were made for each age group separately. A complete set of the self-description items is provided in Appendix B.

Design and Procedure

Each child participated in three experimental sessions, in a fixed order, on consecutive days (with the exception of four children who participated in two sessions on the same day). In the first session, where each child was seen individually by the experimenter, the judgement task was completed. In the second session, children completed the self-description task alone with no other children present in the experimental room. In the third session, children completed the self-description task in groups of 3 to 5 same-sex peers (with the exception of one group of two boys). The group condition was placed in the final session to avoid contamination of both the judgement task and the individual self-description. All boys were tested by a male experimenter, and all girls were tested by a female experimenter.

Results

For responses on both tasks, children scored 0 for selecting 'No', 1 for 'Sometimes', and 2 for 'A lot'.

Judgement Task

We first confirmed that the younger children's stereotypes were indeed more rigid than the older children's stereotypes. A composite score of the rigidity of stereotyping was obtained based on the following reasoning. Children whose stereotypes are rigid should not rate the toy preferences of stereotypical and counterstereotypical target characters differently. On the other hand, children whose stereotypes are less rigid should see that the toy preferences of a stereotypical boy will be more masculine (difference #1) and less feminine (difference #2) than the toy preferences of a counterstereotypical boy. Similarly, they should also understand that the toy preferences of a stereotypical girl will be more feminine (difference #3) and less masculine (difference #4) than the toy preferences of a counterstereotypical girl.

The following steps were used to calculate the composite score. First, for each stereotypical or counterstereotypical target character, the participating children received a 'masculine score' for the two masculine items and a 'feminine score' for the two feminine items (both out of four). The four differences mentioned above were then calculated and summed. This total was taken as a composite score of stereotype rigidity, where lower scores indicate more rigidity and higher scores indicate more flexibility.

A two-way analysis of variance on these composite scores was then conducted, using **age group** and **sex** as the between-subjects variables. This showed a significant main effect of **age group** ($F(2, 58) = 6.70$, $p = .002$) and a significant main effect of **sex** ($F(1, 58) = 17.92$, $p < .001$). Planned contrasts showed that, in confirmation of our hypothesis, the 4- and 6-year-olds scored significantly lower (i.e. were more rigid)

than the 8-year-olds (means: 3.48, 2.38, and 6.86, with ascending age, $p < .001$), while the mean scores of the 4-year-olds and 6-year-olds did not significantly differ. In addition, boys' stereotypes were markedly more rigid than those of girls (means: 2.03 and 6.39, respectively). There was no interaction of **age group and sex** ($F < 0.2$); that is, the effect of age was equivalent for boys and girls.

Self-description Task

Having confirmed that the 4- to 6-year-olds do indeed have more rigid stereotypes than the 8-year-olds, we were able to test our hypothesis that the younger children would be more likely to present themselves as gender-typed in front of a peer audience. For both conditions of the self-description task, each child received a score for the masculine items and a score for the feminine items (both out of 10). Masculine and feminine scores were examined independently in accordance with research findings showing that masculinity and femininity are independent constructs (e.g., Bem, 1974). Due to heterogeneity of variance, arcsine transformations were performed on all scores. An analysis of variance was conducted for sex-typed scores (i.e. masculine scores for the males, feminine scores for the females) and for non-sex-typed scores (i.e. feminine scores for the males, masculine scores for the females). For both analyses, the within-subjects variable was **condition** (alone vs. group) and the between-subjects variables were **age** and **sex**.

First, the analysis on the sex-typed scores showed a significant main effect of **age** ($F(2, 58) = 13.63$, $p < .001$). Reverse Helmert contrasts revealed that the oldest group scored significantly lower than the two younger groups ($p < .0001$). There was also a significant main effect of **condition** ($F(1, 58) = 13.78$, $p < .001$), but this was qualified by a significant interaction of **age and condition** ($F(2, 58) = 3.09$, $p = .05$). Reverse Helmert contrasts revealed that this was due to the fact that only the two younger groups scored higher in the group condition than in the alone condition; the older group showed no change ($p < .03$). No differences were found between the two younger groups. There was also a significant interaction of **sex and condition** ($F(1, 58) = 6.54$, $p < .02$), explained by the fact that boys' scores increased in the group condition while girls' scores did not change. Finally, there was a marginal interaction of **age, sex, and condition** ($F(2, 58) = 2.45$, $p < .10$). Reverse Helmert contrasts revealed that the age and condition interaction observed above occurred only for the males ($p < .05$). In fact, none of the female groups showed much change at all from the alone to the group condition. These patterns are shown in Figure 1. It is worth noting that the observed increase in sex-typed scores for the younger boys was not the result of an undue influence of a few extreme cases: the rounded mean, median, and modal change scores were identical for these children.

For the non-sex-typed scores, the analysis showed that despite the transformations, there was still heterogeneity of variance on the alone condition scores. This seems to be because of the low variance shown by the two older age groups on these scores. The results of this analysis are still given below, but the effects should be interpreted with caution. First, there was a main effect of **age** ($F(2, 58) = 11.09$, $p < .001$). Reverse Helmert contrasts revealed that this is accounted for primarily by the slightly higher scores of the youngest group across both conditions, compared to the older children ($p < .0002$). There was also a significant interaction of **age and sex** ($F(2, 58) = 3.36$, $p < .05$). Reverse Helmert contrasts (all ps $< .10$) show that while males tend to show a steady decrease with age (means across both conditions, 2.45, 1.40, 0.15), females

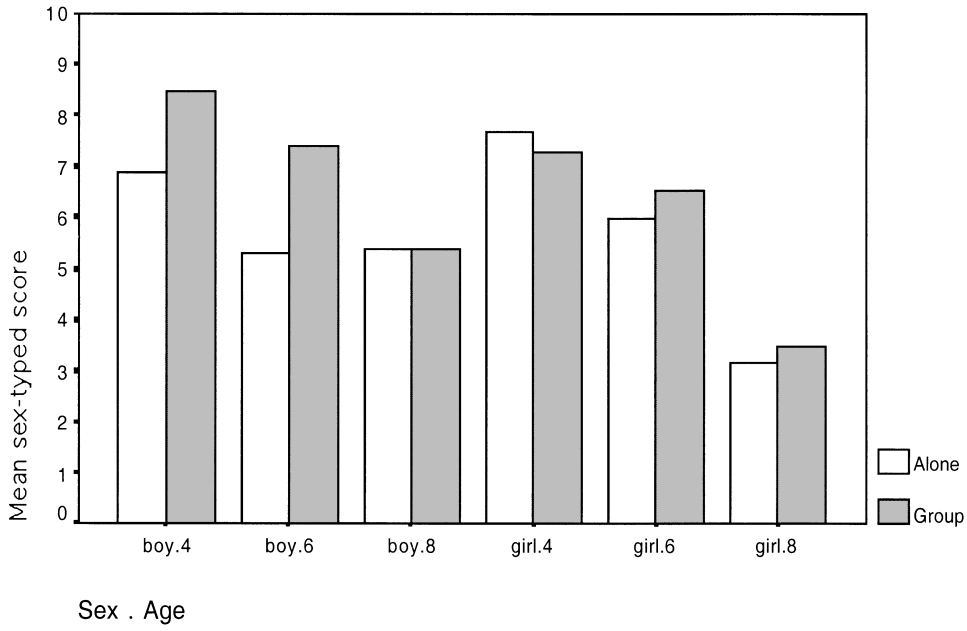

Figure 1. Mean sex-typed scores for self-description task (out of 10), by condition, age, and sex.

show a strong dip in scores in the middle age group followed by an increase in the oldest group (means across both conditions, 3.45, 0.50, 1.75). Finally, there was a main effect of **condition** ($F(1, 58) = 4.94$, $p = .03$). Rather anomalously, scores were generally slightly higher in the group condition compared to the alone condition (overall means, 1.86 vs. 1.38, respectively). There were no interactions of the **condition** variable with **age** or **sex** (all Fs < 1). It is worth noting additionally that the 4- to 6-year-old boys' increase in non-sex-typed scores from the alone to the group condition was significantly smaller than their increase in sex-typed scores (mean increases 1.81 for sex-typed scores and .86 for non-sex-typed scores; $t(20) = 2.42$, $p = .01$).

Finally, in confirmation of our hypothesis, the extent to which children increased their sex-typed scores from the alone to the group condition was found to be correlated with the scores on the judgement task ($r(62) = -.23$, $p < .05$). That is, the less flexible the children's stereotypes, the more the children tried to appear sex-typed in front of the peer audience. In contrast, the change in non-sex-typed scores was not significantly correlated with scores on the judgement task ($r(62) = -.09$, *ns*).

We followed up the initial correlation between the change in sex-typed scores and the scores on the judgement task, by recalculating the coefficient after controlling for the effects of gender and age group. Initially, this partial correlation was not significant (partial $r(60) = -.12$, $p = .17$), with an unexpectedly low correlation within the female sample. However, several of the girls scored very highly on the judgement task, and this presented a problem for our tests of linear association. Specifically, simple regression analysis, fitting a straight line to the data, estimated that scores above 10 on the judgement task will be associated with *decreases* in sex-typed scores from the alone to the group condition. In fact, the ten children (all girls of varying ages) who scored above 10 on the judgement task showed a negligible mean increase in sex-typed scores of 0.60. This is consistent with our theoretical account, which predicts that high

scores on the judgement task should be associated with *no difference* between alone and group conditions. We therefore excluded the ten girls who scored above 10 on the judgement task, and recalculated the coefficients for the correlation between the judgement task and the increase in sex-typed scores from the alone to the group condition. The correlation was then observed both within the male sample and within the female sample (rs = −.27 and −.34, respectively). Thus, after controlling for both gender and age group, the partial correlation was statistically significant (partial $r(50) = -.26$, $p < .05$).

Discussion

Our hypothesis that children with more rigid gender stereotypes would be more motivated to present themselves as sex-typed to their peers was supported by this experiment. Children aged as young as 4 do appear to be responsive to social evaluation in domains where they are highly motivated to make good impressions. In fact, in this experiment, it was the oldest age group (the 8- to 9-year-olds) who were not influenced by the presence of their peer group to appear more sex-typed. This highlights the significance of considering domain differences, given findings from previous research showing that social evaluation tends to become increasingly salient from ages 8 to 9 (e.g., Banerjee & Yuill, 1999a; Vasey, Crnic, & Carter, 1994).

As predicted, differences in the rigidity of gender stereotypes are associated with the extent to which the presence of an audience elicits more sex-typed self-descriptions. This was found to be true not only in terms of differences between individuals (the correlational analysis) but also in terms of differences between groups of children. In particular, the younger boys, who showed the most rigid gender stereotypes, were more motivated than the older boys and the girls in all age groups to try and appear sex-typed in front of a same-sex peer audience. It is interesting to note that the peer audience had little effect on the self-descriptions of the younger girls. These children, while more rigid in their judgements than the oldest girls, were still considerably more flexible in comparison with the younger boys. Correspondingly, they were less motivated to adjust their self-descriptions to appear more gender-typed in front of the peer audience. This pattern of results is not entirely surprising, in view of several studies which suggest that the gender role of females does not demand the strict adherence that the male gender role does. For example, play with other-sex toys in general is higher in female than in male preschoolers (Serbin et al., 1979), and several researchers have noted that gender-typing is not as strong for girls as it is for boys (e.g., see Huston, 1985). The fact that we observed little evidence of a motivation among the girls to adjust their self-descriptions to appear more gender-typed is therefore consistent with findings in past research and with our argument that such a motivation is linked closely to rigidity in gender-typing.

It should be noted that the group differences between males and females observed in this study are a little ambiguous since the gender of the participant always matched the gender of the peer-group audience. The correspondence between the change in sex-typed scores in the self-description task and the flexibility scores in the judgement task suggests that there are genuine differences in the orientations of boys and girls. However, one could speculate that compounding this difference in the present study is a tendency for male peer audiences in general to elicit more conformity to gender role norms than female peer audiences. This possibility deserves attention in future research.

It is useful to consider our findings in light of Serbin et al.'s (1979) study of preschoolers' play behaviour, mentioned earlier. In that study, both boys and girls were found to play less with other-sex toys when in front of peers of the other sex than when alone. This raises two interesting issues. First, the link between children's self-descriptions and their actual play in everyday social interactions needs to be explored. Real-life play with peers involves a web of actions, reactions, and evaluations that is obviously more complex than the structured self-description task used here. Determining exactly how the motivations examined in the present study map onto the dynamics of real-life play with peers is an important topic for future research. Second, following on from the suggestion in the preceding paragraph, it would seem to be instructive to consider these issues in the context of interactions with peers of both the same and the other sex.

Overall, our results are consistent with the existing research on the development of gender schemas. As noted earlier, the gender schematic processing model (see Martin, 1989; Martin & Halverson, 1981) makes the point that children's gender stereotypes become increasingly flexible with age. This evidently translates into real-life social behaviour. Although older children clearly do not describe themselves as typical of the other sex, they are less likely than younger children to describe themselves as highly sex-typed. Importantly, when social evaluation concerns are high, these general tendencies are amplified in front of an audience. Thus, the young boys, for whom sex-typing is high in the baseline condition, are even more sex-typed in their self-descriptions when same-sex peers are present. This pattern is compatible with suggestions that a large part of the concern about social evaluation is a desire to appear congruent with the norms of the peer group (see Erwin, 1993). In the case of younger boys, norms for gender-related behaviours are extremely strong and the self-presentational concerns in that domain therefore acquire corresponding strength.

The one anomalous finding is that the non-sex-typed scores, while still very low on average, showed a slight increase in the group condition. This is unexpected since we would expect the presence of an audience to amplify the general tendency for low non-sex-typed scores. However, it is possible that with this task, children had a general tendency to say that they did more of everything in front of an audience. This kind of artifactual explanation is compatible with the finding that this change in non-sex-typed scores was *not* correlated with the rigidity of gender stereotypes, unlike the change in sex-typed scores.

The findings of the present study contribute to our awareness of self-presentational motives in gender-related behaviour, as discussed above. Importantly, they also inform our theoretical understanding of how self-presentation in general develops. As explained earlier, although social evaluation may assume particular importance in many domains from age 8 onwards, the results demonstrate that concern about one's public image is likely to be a salient feature in at least some of young children's interactions. This is consistent with previous findings that children have the cognitive capacity for understanding self-presentational motives when they start school (Banerjee & Yuill, 1999b). The present study presents a valuable insight into young children's capacity for sophisticated self-regulatory responses to social situations that involve interpersonal evaluations.

In conclusion, these findings highlight for the first time the important point that children's cognition about gender-related behaviour has an impact on their self-presentational strategies. Children are not passively shaped by peer evaluation; instead, when sufficiently motivated, they will actively use their knowledge about gender

stereotypes to manipulate that evaluation. As children grow older, their view of gender stereotypes becomes less rigid, and social evaluation concerns appear in the context of peer group norms in other domains (see Banerjee & Yuill, 1999a; Fine, 1988). We suggest that researchers in the area of gender development must now address the distinction between private preference and public performance and consider how the two relate to each other (e.g., see Baumeister, 1986; Snyder, 1987). The present research has demonstrated that such a distinction is significant in young boys' gender-related behaviour and that this is related to their rigid gender stereotypes. However, there are individual differences in children's cognition about gender-related behaviour (see Levy & Carter, 1989; Signorella, 1987), and given the present findings it seems likely that these individual differences will have consequences for social behaviour.

A major goal for future research must be to determine the specific origins of these differences in the use of gender-related behaviour to shape social evaluation. In seeking to explain the age effects observed here, one can point to cognitive constraints as making younger children more rigid in their gender-role orientations and therefore more motivated to appear gender-typed in public situations. However, the existence of gender differences and indeed of individual differences within genders and age groups tells us that the cognitive characteristics involved are likely to be coupled with social-motivational influences that arise from idiosyncratic experiences and from society at large. The task remains for researchers to formulate an understanding of how these various factors interact with each other to generate differences in children's gender-related judgements and behaviours.

References

Banerjee, R. & Yuill, N. (1999a). Children's explanations for self-presentational behaviour. *European Journal of Social Psychology*, *29*, 105–111.

Banerjee, R. & Yuill, N. (1999b). Children's understanding of self-presentational display rules: Associations with mental-state understanding. *British Journal of Developmental Psychology*, *17*, 111–124.

Baumeister, R. F. (Ed.). (1986). *Public self and private self*. New York: Springer-Verlag.

Bem, S. L. (1974). The measurement of psychological androgyny. *Journal of Consulting and Clinical Psychology*, *42*, 155–162.

Bigler, R. S. & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, *63*, 1351–1363.

De Lucia, L. A. (1963). The Toy Preference Test: A measure of sex role identification. *Child Development*, *34*, 107–117.

Erwin, P. (1993). *Friendship and peer relations in children.* Chichester: Wiley.

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

Fagot, B. I. (1985). Beyond the reinforcement principle: Another step toward understanding sex role development. *Developmental Psychology*, *21*, 1097–1104.

Fine, G. A. (1988). Friends, impression management, and preadolescent behaviour. In G. Handel (Ed.), *Childhood socialization* (pp. 209–233). New York: Aldine de Gruyter.

Golombok, S. & Fivush, R. (1994). *Gender development*. Cambridge, UK: Cambridge University Press.

Golombok, S. & Rust, J. (1993). The measurement of gender role behaviour in pre-school children: A research note. *Journal of Child Psychology and Psychiatry*, *34*, 805–811.

Huston, A. C. (1985). The development of sex typing: Themes from recent research. *Developmental Review*, *5*, 1–17.

Lamb, M. E., Easterbrooks, A., & Holden, G. W. (1980). Reinforcement and punishment among preschoolers: Characteristics, effects and correlates. *Child Development*, *51*, 1230–1236.

Lamb, M. E. & Roopnarine, J. L. (1979). Peer influences on sex role development in preschoolers. *Child Development*, *50*, 1219–1222.
Levy, G. D. & Carter, D. B. (1989). Gender schema, gender constancy, and gender-role knowledge: The roles of cognitive factors in preschoolers' gender-role stereotype attributions. *Developmental Psychology*, *25*, 444–489.
Martin, C. L. (1989). Children's use of gender-related information in making social judgments. *Developmental Psychology*, *25*, 80–88.
Martin, C. L. & Halverson, C. (1981). A schematic processing model of sex typing and stereotyping in children. *Child Development*, *52*, 1119–1134.
Powlishta, K. K., Serbin, L. A., Doyle, A.-B., & White, D. R. (1994). Gender, ethnic, and body type biases: The generality of prejudice in childhood. *Developmental Psychology*, *30*, 526–536.
Rubin, J. Z., Provenzano, F. J., & Luria, Z. (1974). The eye of the beholder: Parents' views on sex of newborns. *American Journal of Orthopsychiatry*, *44*, 512–519.
Seavey, A. A., Katz, P. A., & Zalk, S. R. (1975). Baby X: The effect of gender labels on adult responses to infants. *Sex Roles*, *1*, 103–109.
Serbin, L. A., Connor, J. M., Burchardt, C. J., & Citron, C. C. (1979). Effects of peer presence on sex-typing of children's play behavior. *Journal of Experimental Child Psychology*, *27*, 303–309.
Signorella, M. L. (1987). Gender schemata: Individual differences and context effects. In L. S. Liben & M. L. Signorella (Eds.), *Children's gender schemata: New directions in child development (No. 38, pp. 5–22)*. San Francisco: Jossey-Bass.
Snyder, M. (1987). *Public appearances/private realities: The psychology of self-monitoring.* New York: Freeman.
Urberg, K. A. (1982). The development of the concepts of masculinity and femininity in young children. *Sex Roles*, *8*, 659–668.
Vasey, M. W., Crnic, K. A., & Carter (1994). Worry in childhood: A developmental perspective. *Cognitive Therapy and Research*, *18*, 529–549.

Acknowledgements

We would like to extend our thanks and appreciation to the children and staff at the school that participated in this study. We are grateful for the useful comments of Clara Strauss, Nicola Yuill, and three anonymous reviewers on earlier versions of this paper. This research was funded by Economic and Social Research Council (UK) Research Studentship No. R00429534012.

Appendix A: Judgement Task

1. This is a boy called Tom. His best friend is a boy, and he likes to play with aeroplanes.

Do you think Tom would like to:

play with dolls?	NO	SOMETIMES	A LOT
play football?	NO	SOMETIMES	A LOT
do skipping?	NO	SOMETIMES	A LOT
play with toy guns?	NO	SOMETIMES	A LOT

2. This is a boy called John. His best friend is a girl, and he likes playing with prams/buggies.

Do you think John would like to:

play with dolls?	NO	SOMETIMES	A LOT
play football?	NO	SOMETIMES	A LOT
do skipping?	NO	SOMETIMES	A LOT
play with toy guns?	NO	SOMETIMES	A LOT

3. This is a girl called Julie. Her best friend is a girl, and she likes to play kitchens. Do you think Julie would like to:

play with dolls?	NO	SOMETIMES	A LOT
play football?	NO	SOMETIMES	A LOT
do skipping?	NO	SOMETIMES	A LOT
play with toy guns?	NO	SOMETIMES	A LOT

4. This is a girl called Sarah. Her best friend is a boy, and she likes to play with cars. Do you think Sarah would like to:

play with dolls?	NO	SOMETIMES	A LOT
play football?	NO	SOMETIMES	A LOT
do skipping?	NO	SOMETIMES	A LOT
play with toy guns?	NO	SOMETIMES	A LOT

Note: A pictorial scale of a cross, one tick, and two ticks was used to represent the three response options.

Appendix B: Self-description Items

1. Do you ride bikes? (N)	NO	SOMETIMES	A LOT
2. Do you play football? (M)	NO	SOMETIMES	A LOT
3. Do you do skipping? (F)	NO	SOMETIMES	A LOT
4. Do you play with dolls? (F)	NO	SOMETIMES	A LOT
5. Do you play with cars? (M)	NO	SOMETIMES	A LOT
6. Do you play kitchens? (F)	NO	SOMETIMES	A LOT
7. Do you watch the television? (N)	NO	SOMETIMES	A LOT
8. Do you play with aeroplanes? (M)	NO	SOMETIMES	A LOT
9. Do you do wrestling and fighting? (M)	NO	SOMETIMES	A LOT
10. Do you play with toy buggies? (F)	NO	SOMETIMES	A LOT
11. Do you play with toy guns? (M)	NO	SOMETIMES	A LOT
12. Do you play with makeup and clothes? (F)	NO	SOMETIMES	A LOT

N = neutral item
M = masculine item
F = feminine item

Note: A pictorial scale of a cross, one tick, and two ticks was used to represent the three response options.