Developmental Science

Developmental Science 13:2 (2010), pp 298–306 DOI: 10.1111/j.1467-7687.2009.00880.x

PAPER

The development of spontaneous gender stereotyping in childhood: relations to stereotype knowledge and stereotype flexibility

Rainer Banse,[1] Bertram Gawronski,[2] Christine Rebetez,[3] Hélène Gutt[4] and J. Bruce Morton[2]

1. *Department of Psychology, University of Bonn, Germany*
2. *Department of Psychology, University of Western Ontario, London, Canada*
3. *Department of Psychology, University of York, UK*
4. *Faculté des Sciences Psychologiques et del l'Education, Université Libre de Bruxelles, Belgium*

Abstract

The development of spontaneous gender stereotyping in children was investigated using the newly developed Action Interference Paradigm (AIP). This task consists of assigning gender-stereotypical toys as quickly as possible to boys and girls in either a stereotype-congruent or a stereotype-incongruent manner. A pilot study with 38 children (mean age 5.1 years) provided evidence for spontaneous gender stereotyping in the AIP, which was reflected in higher latencies for stereotype-incongruent compared with stereotype-congruent toy assignments. The main study, with 66 children (aged 5, 8 and 11 years), compared the development of spontaneous stereotyping with established measures of stereotype flexibility and stereotype knowledge. Stereotype flexibility showed a strong increase from age 5 to 11. In contrast, stereotype knowledge and spontaneous stereotyping remained stable at high levels. The results provide evidence for a dissociation between stereotype flexibility and spontaneous stereotyping, suggesting that spontaneous stereotyping may be more closely related to stereotype knowledge than to stereotype flexibility.

Introduction

The last two decades of research on social stereotyping have been characterized by an increasing interest in spontaneous or automatic manifestations of stereotypes and their relation to more controlled and deliberate information processing. This interest is motivated by empirical demonstrations of dissociations between the knowledge of versus the belief in a social stereotype. For example, an individual may know that most people think that a specific social group is less intelligent than another group, and simultaneously refute this stereotype as wrong. According to Devine's (1989) influential dissociation model of prejudice and stereotyping, stereotype knowledge is often activated automatically, whereas the rejection of this knowledge requires controlled, cognitively effortful processes (see Bargh, 1994). Such dissociations are also reflected in various dual-process models of social cognition, which postulate two general modes of information processing (e.g. Gawronski & Bodenhausen, 2006; Smith & DeCoster, 2000; Strack & Deutsch, 2004). In a nutshell, these models distinguish between (a) the activation of associative knowledge structures in memory, which can occur independently of whether a person considers these associations as accurate or inaccurate, and (b) the propositional validation of activated information as either true or false, which serves as the basis for deliberate decisions and action intentions. These dissociations raise important questions concerning the organization of cognitive processes underlying overt behaviour. In line with this concern, automatically activated associations have been shown to influence spontaneous behaviour even when these associations are deliberately rejected as inaccurate or false (for a review, see Strack & Deutsch, 2004).

The study of developmental change in spontaneous stereotyping is another potentially fruitful means of gaining insight into the underlying organization of social cognition. It is generally assumed that stereotype knowledge is acquired early in childhood, is highly overlearned, and is relatively resistant to change (e.g. Devine, 1989; Wilson, Lindsey & Schooler, 2000). By the age of three years, children readily distinguish between males and females, and associate certain objects (e.g. trucks) more strongly with one gender (boys) than with the other. At the same time, with development, children typically show a strong increase in gender stereotype

Address for correspondence: Rainer Banse, Department of Psychology, University of Bonn, Kaiser-Karl-Ring 9, 53111 Bonn, Germany; e-mail: banse@uni-bonn.de

flexibility and recognition that the gender stereotype can be inaccurate (e.g. Serbin & Sprafkin, 1986).

For instance, older children admit that although trucks are more commonly associated with boys, girls can play with trucks as well. Thus, to the degree that stereotype knowledge can influence spontaneous behaviour (Strack & Deutsch, 2004), the emergence of spontaneous discriminatory behaviour in childhood may resemble the developmental course of stereotype knowledge acquisition. In other words, even when children have acquired the mental flexibility to reject a social stereotype, this newly acquired flexibility may leave behavioural manifestations of automatically activated stereotypes unaffected (see Greenwald & Nosek, 2008).

Despite the large body of research on the development of gender stereotype knowledge and gender stereotype flexibility, little is known about the development of spontaneous gender stereotyping in children. Following the literature on stereotyping in social psychology, we use the term *spontaneous stereotyping* for behavioural manifestations of automatically activated social stereotypes under conditions of little cognitive control (e.g. owing to a lack of cognitive resources, attention, or time). The concept of stereotyping is akin to the concept of *social discrimination*, but less normatively laden, and hence more appropriate in the context of gender stereotyping in children.

Documenting developmental changes in spontaneous stereotyping and clarifying whether they converge with developmental changes in stereotype knowledge or stereotype flexibility provides important insights into the determinants of social stereotyping in children. In addition, research on these questions can answer important theoretical questions concerning the organization of cognitive processes underlying social stereotypes (e.g. Devine, 1989). The present study addressed these questions through the use of a new latency-based measure of spontaneous stereotyping, the *Action Interference Paradigm* (AIP). Suitability of the AIP for assessing spontaneous gender stereotyping in children was assessed by means of a pilot study involving children aged 3 to 6 years. Based on these findings, the main study was conducted to investigate the convergence versus divergence of the developmental courses of spontaneous gender stereotyping, stereotype flexibility, and stereotype knowledge in children aged 5 to 11 years.

The development of gender stereotypes in childhood

The ability to distinguish between males and females not only contributes to the development of sexual identity, but represents a core category of knowledge about the social world. According to Diekman and Eagly (2000), the gender stereotype of a specific culture typically reflects the predominant gender roles of this culture. These role-congruent stereotypes often extend to non-social objects by means of the differential use of particular objects by males and females, such as the use of different kinds of tools by women and men (e.g. iron, hammer) or of toys by girls and boys (e.g. dolls, trucks).

Gender-related knowledge develops early in childhood. By 3 years, for example, children can correctly classify photographs of men and women in terms of their gender (Thompson, 1975) and possess basic knowledge of gender stereotypes. In one study (Edelbrock & Sugawara, 1978), children aged 3 to 5 years were presented with pictures showing gendered activities of adults or children (e.g. adult holding a baby, child holding a baseball club) and were asked whether the activities were for boys, for girls, or both boys and girls. This task, termed the Sex Role Learning Index (SERLI), revealed high levels of stereotypical knowledge among children aged 3 to 4.5 years, with further increases between 4.5 and 5 years. In a study of 3- to 7-year-olds, Serbin and Sprafkin (1986) used the SERLI to examine age-related changes in stereotype knowledge and stereotype flexibility, which was operationalized as the use of the 'both' category in the task. Whereas gender stereotype knowledge showed a marked increase between 3 and 7 years, gender stereotype flexibility decreased until the age of 6 years before it started to increase between the ages of 6 and 7 years.

This result was confirmed in a meta-analysis by Signorella, Bigler and Liben (1993), who found an increase of gender stereotype knowledge from age 3 to 7 years in 45 studies using a forced-choice male–female answer format. In a sample of 54 studies that also included a 'both' answer category, they found an initial decrease of stereotype flexibility until primary school age, followed by an increase during primary school until the age of 10 years. This increase of gender stereotype flexibility was also demonstrated in a longitudinal study by Trautner *et al*. (2005). Interestingly, this study also provided evidence that individual differences in stereotype rigidity/flexibility in children aged 5 to 7 were unrelated to individual differences at age 10. However, children who showed an early peak of stereotype rigidity showed an earlier onset of increasing flexibility, whereas children with a later peak rigidity had a later onset of increasing flexibility, and reached a lower level of flexibility at age 10. These results suggest that all children have to go through the same developmental stages of gender stereotype acquisition, with a peak of rigid stereotype application, followed by increasing stereotype flexibility. Moreover, individual differences within age groups resulted from children undergoing the very same U-shaped developmental trajectory earlier or later.

In summary, the available evidence suggests that children have a basic notion of the gender stereotype of their culture as early as the age of 3 years. This knowledge strongly increases until reaching ceiling level at the age of approximately 7 years. In contrast, the development of gender stereotype flexibility follows a U-shaped trajectory. Starting with the acquisition of the

gender stereotype at the age of 2 years, children use this knowledge in an increasingly rigid manner, with flexibility reaching the lowest level between 5 and 7 years. After that, the use of gender stereotypes becomes increasingly flexible again, with stereotype flexibility reaching ceiling levels around the ages of 10 to 12 years.

Spontaneous gender stereotyping in children

The effects of spontaneous gender stereotyping in children have been studied in various behavioural domains. Widen and Russell (2002) demonstrated that preschoolers perceived a given emotional expression in a stereotype-consistent way depending on the gendered hair style of the adolescent target person. Bennett and Sani (2008) provided evidence for spontaneous self-stereotyping in 7-year-olds, who described themselves as more similar to same-sex peers if gender was made salient than they did in a control condition.

Most, Sorber and Cunningham (2007) used an auditory Stroop task to assess spontaneous gender stereotyping in children aged 8 to 10 years. In this task, female and electronically modified pseudo-male versions of the same utterances pronounced female and male names (e.g. Amy, Brian) or stereotypically female or male words (e.g. cheerleader, baseball). The task of the participants consisted of classifying the gender of the speaker by pronouncing the word 'boy' or 'girl' (Experiment 1 with adults), or by pressing a response key labelled 'boy' or 'girl' (Experiment 2 with children). As with the classic Stroop effect, it was expected that the gender discrimination task would be facilitated if the task-irrelevant meaning of the word was stereotypically congruent with the gender of the speaker, whereas stereotypically incongruent words should lead to interference. An interesting feature of this paradigm is the comparison of unambiguously male and female first names and stereotypically male and female words, which provides a means to evaluate the strength of interference of stereotypically gendered words. For the latter, interference is not obligatory, but is a function of the strength of the gender stereotype. As expected, both 8-year-old children and adults showed a significant Stroop interference effect for both names and gender-stereotypical words. Interestingly, in the adult sample this interference effect was stronger for gender-typical first names, whereas for the children interference was stronger for gender-stereotypical words. Although not conclusive, these data suggest that gender stereotypes may be so salient to 8-year-olds that gendered words have a stronger impact than gender-typical names.

The present research

The present research expands on previous work by investigating behavioural manifestations of spontaneous gender stereotyping in children by using a newly developed behavioural paradigm: the Action Interference Paradigm (AIP). The logic of the AIP is similar to that of the Stroop variant employed by Most *et al*. (2007), in that the AIP is based on the potential interference between an unintended, prepotent response tendency (i.e. a spontaneous tendency to show a stereotype-congruent response) and the accurate response required in the task (i.e. a stereotype-incongruent response). So that the task is suitable for very young children, the AIP does not involve written stimuli or labels, but only photographs of toys that are stereotypically preferred by girls (e.g. dolls) or boys (e.g. trucks). These toys have to be assigned to girls or boys as quickly as possible.

In order for spontaneous manifestations of gender stereotyping to be assessed, each child completed the toy assignment task in a stereotype-congruent (giving stereotypically male toys to a boy and stereotypically female toys to a girl) and in a stereotype-incongruent (giving stereotypically male toys to a girl and stereotypically female toys to a boy) manner. The assumption was that spontaneous gender-stereotypical action tendencies (e.g. to give a doll to the girl) should facilitate responses when the required response was stereotype-congruent, but interfere with responses when the required response was stereotype-incongruent (i.e. give a doll to the boy). To render the experimental paradigm meaningful and involving for children, the task was presented as a game in which children were asked to help Santa Claus quickly distribute Christmas presents to other children. There were 60 trials in both the stereotype-congruent and the stereotype-incongruent task block. In a pilot study, we tested the suitability of the AIP for children aged 3 to 6 years. In the main study, we used the AIP to investigate the convergence versus divergence of the developmental courses of spontaneous gender stereotyping, stereotype flexibility, and stereotype knowledge in children aged 5 to 11 years.

Pilot study

To test whether the AIP is suitable for assessing spontaneous gender stereotyping in preschool children, a pilot study was run with children attending a dedicated observation lab preschool at the University of Western Ontario, Canada. Out of 44 participating children, three had difficulties following the instructions, and three had to be excluded from data analysis because of error rates in excess of 30% in at least one block of 30 trials. The remaining sample of 38 children (18 boys, 20 girls), with a mean age of 5.1 years (ranging from 3.5 to 6.3), successfully completed the AIP with error rates between 5% and 6%. To familiarize children with the assignment task, 40 practice trials with red and blue parcels preceded the AIP. The order of the stereotype-congruent and stereotype-incongruent blocks was counterbalanced. In addition, we fully counterbalanced the particular key assignments in the task.

There were three important results. First, responses were slower in the stereotype-incongruent than in the stereotype-congruent assignment task, indicating that the AIP indeed captured the expected stereotype interference effect. Second, error rates were significantly influenced by the order of the two blocks (i.e. they were significantly higher in the second block after changing the key assignment), but not by stereotype congruency per se. Third, although the expected interference effect was obtained in both the congruent–incongruent and the incongruent–congruent order condition, it tended to be stronger when the stereotype-congruent task was administered before the stereotype-incongruent task. These results suggest that, for optimizing the AIP as a measure of spontaneous stereotyping, the score should be based on response latencies (not errors), and the critical blocks should be presented in the congruent–incongruent order (and not the reverse). In the following main study, the AIP was set up and scored accordingly.

Main study

The main study was conducted to investigate how spontaneous manifestations of gender stereotyping in the AIP relate to the development of stereotype knowledge and stereotype flexibility in children aged 5, 8 and 11 years. As mentioned earlier, children acquire gender stereotype knowledge between 2 and 6 years of age, and tend to apply this knowledge in an increasingly rigid manner during this period (Trautner *et al*., 2005). Beginning at about the age of 6 years, however, the use of stereotypical knowledge becomes increasingly flexible, until stereotype flexibility reaches ceiling levels at age 10 to 12 years.

The present study focused on three groups of children aged 5, 8 and 11 years. According to the longitudinal data of Trautner *et al*. (2005), this age range covers the period of maximal change of gender stereotype flexibility, from peak stereotype rigidity (between 5 and 7 years) to high stereotype flexibility (at age 10), with gender stereotype knowledge approaching ceiling levels at age 8.

Drawing on these findings, we expected that even the youngest children in our sample would show high levels of gender stereotype knowledge, with some margin for increases between ages 5 and 8. Stereotype flexibility, in contrast, was expected to show a marked age-related increase across the three groups, as reflected in a more frequent use of the 'both' category in the SERLI when children are asked which gender uses stereotype-related toys and objects.

For spontaneous gender stereotyping, we expected children to show high levels of action interference in the AIP, irrespective of the particular age group. This prediction is based on the dual-process assumption that activated knowledge can influence spontaneous responses even if this knowledge contradicts personally held beliefs (Devine, 1989; Strack & Deutsch, 2004). Moreover, age-related increases in stereotype flexibility should be insufficient to qualify the impact of stereotype knowledge on spontaneous stereotyping (Devine, 1989).

Method

Participants

The participants were 66 children, all attending the same school in a suburb of Brussels, Belgium. For all participating children, written, informed consent of the parents was obtained prior to the experiment. Two children had to be excluded because they did not follow the task instructions. The data of two children were discarded because of high error rates (> 30%) in at least one trial block. The error rate of the remaining sample was 7.9%. The remaining sample of 33 girls and 29 boys included 20 5-year-olds (10 girls) of mean age 5.4 years (*SD* = .34, range = 4.5 to 5.8), 19 8-year-olds (11 girls) of mean age 7.5 years (*SD* = .38, range = 6.8 to 8.4), and 23 11-year-olds (12 girls) of mean age 10.6 years (*SD* = .25, range = 10.2 to 11.2). The exact ages of the children were obtained from school files.

Materials

For assessing gender stereotype knowledge and gender stereotype flexibility, two sets of pictures were prepared on 10 × 10 cm cards. The first set was taken from the Sex Role Discrimination scale (SRD; Edelbrock & Sugawara, 1978) and comprised 22 black and white drawings of common objects, 10 of which were stereotypically associated with women (e.g. iron, vacuum cleaner, cooking utensils) and 10 with men (e.g. hammer and nails, helmet, rifle), as well as two neutral objects (bed, ice cream). All the original items were used except a baseball bat and ball, which were replaced by soccer shoes and a soccer ball, as those are more common in Europe. A second set of 30 colour photographs showed toys that previous studies found were stereotypically preferred by either boys (e.g. airplane, racing car, garage) or girls (e.g. doll, doll's house, doll's clothing; Carter & Patterson, 1982; Connor & Serbin, 1977; Eisenberg, Murray & Hite, 1982; Fagot, 1985; Liss, 1981; O'Brian & Huston, 1985; Richardson & Simpson, 1982). Photographs were digitized and used as stimuli for the AIP. For a practice phase, pictures of a blue and a red gift box were used.

Procedure and measures

To avoid problems of fatigue[1], the main experiment was run in two sessions. The measures of stereotype knowledge and stereotype flexibility were obtained in the morning, and the measure of spontaneous

[1] Pre-tests with 15 children, including 9 children aged 5 years, revealed that the younger children had problems concentrating towards the end of a single session.

stereotyping in the afternoon of the same day for each child. This order had the advantage that the children were familiarized with the toy photographs during the morning session before encountering the same photos when completing the AIP. All assessments were run in individual sessions in a quiet room.

Stereotype knowledge and stereotype flexibility. After a short introduction to the experiment and the female experimenter in the classroom, each child was accompanied individually to a separate room. Stereotype knowledge and stereotype flexibility were assessed using the instructions developed by Edelbrock and Sugarawa (1978) and the scoring method used by Serbin and Sprafkin (1986). Each child was shown three boxes for items that are used (a) only by women (girls), (b) only by men (boys), or (c) by both women and men (girls and boys), as indicated by corresponding photographs on the boxes. The experimenter showed each card to the child and said: 'This is an [iron]. Who can use this, a mummy, a daddy, or both?' The response alternatives were illustrated by placing the card above the three boxes. Then the card was given to the child to put it in one of the three boxes. This was done for all cards, starting with the common objects followed by the toy cards. The proportion of 'both' classifications was used as an index of stereotype flexibility. After all the cards had been assigned, the answer box 'both' was taken away and the child was asked to classify the cards from this category again: 'Who uses this object more often, a mummy or a daddy (a girl or a boy)?' The cards and answer alternatives were presented in the same randomized order for all children. The total proportion of stereotype-congruent classifications of the combined first and second sorting task was used as an index of gender stereotype knowledge.

Spontaneous gender stereotyping. In the afternoon session, each child was again accompanied to a separate room and seated in front of an IBM-compatible laptop computer. The AIP consisted of two tasks comprising 60 trials each, in which a series of photographs of toys stereotypically associated with boys (e.g. trucks) or girls (e.g. dolls) was sequentially presented on a computer monitor. The toys had to be assigned as quickly as possible to either a girl or a boy by pressing one of two designated keys of a button box. In the first block of 60 trials, the required assignment was congruent with the dominant gender stereotype; in the second block of 60 trials the required assignment was incongruent with the dominant gender stereotype.

The children were asked to play a game in which they were going to help Santa Claus bring presents to other boys and girls. To familiarize children with the procedure, they were told that they would first need to practise the task with blue and red gift boxes. In the practice trials, a picture of a red box was placed on the left key of a button box, and a picture of a blue box on the right key. During the practice trials, a series of 40 pictures showing a red or a blue gift box had to be assigned as quickly as possible by using the correct response button. The experimenter made sure that the children understood the task, and encouraged them to respond quickly while avoiding too many errors.

In the stereotype-congruent AIP block, children were told that they now would come to the house where Sarah and Marc were living. Sarah likes dolls, doll's houses, doll's clothes and a number of other doll's items, whereas Marc prefers cars, trucks and tools. Whenever a present for Sarah appeared on the computer screen, the child was asked to press the left response key, and when a present for Marc appeared the child was asked to press the right key. The response keys were indicated by two small photographs of Sarah and Marc that were fixed directly on the respective response keys, and by two larger photographs on top of the keyboard at the appropriate side. The children were encouraged to respond quickly because Santa Claus had a lot of work to do.

After accomplishing 60 stereotype-congruent trials, the photographs were replaced by a second set. The experimenter introduced Pierre and Isabelle and their gender stereotype-incongruent toy preferences. Dolls, doll's houses and doll's clothing had to be given to Pierre by pressing the left response key; cars, tools and trucks had to be given to Isabelle by pressing the right response key. The procedure of the AIP was identical to that of the pilot study except for the following features. Instead of fully counterbalancing method-factors (i.e. the order of the stereotype-compatible and stereotype-incompatible trial blocks, the key assignments of toy-types, children's pictures and names, changing or not the key assignment across blocks), all factors were kept constant in the main study to reduce systematic error variance. Such error variance may reduce the statistical power to detect potential differences across age groups, which would work in favour of our hypothesis that spontaneous gender stereotyping does not differ across age groups. Moreover, at the beginning of each sub-block of 30 trials, the experimenter stated the names of the children who were supposed to receive the respective toys. This modification was based on the results of a pre-test, in which several (mostly older) children reported that they had tried to ignore the gender of the children in the incongruent task (e.g. by recoding the response alternatives as left–right instead of boy–girl) because they were confused by this information. The names of the boy and the girl were repeated in the middle of the congruent and incongruent task block to discourage a recoding of the task. After finishing the experiment all participants were thanked and later collectively informed about the aims of this research in their classroom.

Results

Stereotype knowledge

An index of gender stereotype knowledge was calculated as the proportion of stereotype-congruent classifications

for the 20 common objects used by adults that are stereotypically associated with females (e.g. iron) or males (e.g. hammer). A second index of stereotype knowledge was based on stereotype-congruent classifications of the 30 toys that are stereotypically preferred by girls (e.g. doll) or boys (e.g. racing car). Parametric statistical tests (ANOVAs) of all indices based on proportions were conducted using arcsin-transformed variables to render their distribution more normal. As expected, gender stereotype knowledge for common objects used by adults was already very high in the youngest group of 5-year-olds (78.5%; see Table 1 for all means and statistical indices). However, stereotype knowledge still increased significantly to 96.5% in 11-year-olds. For toys, even the youngest age group showed ceiling levels of stereotype knowledge (98.5%), thus leaving little room for a significant increase across age groups.

Stereotype flexibility

Two indices of gender stereotype flexibility were calculated as the proportions of 'both women and men' and 'both girls and boys' responses, respectively. Replicating prior research, stereotype flexibility showed a strong and significant increase across the three age groups. For common objects, stereotype flexibility increased from 33.8% in 5-year-olds to 87.8% in 11-year-olds. The same pattern emerged for toys, for which stereotype flexibility increased from 14.5% in 5-year-olds to 77.83% in 11-year-olds. All age group differences were statistically significant.

Spontaneous stereotyping

To analyse action interference effects, latencies lower than 400 ms (1%) and higher than 10,000 ms (0.3%) were discarded because they are conventionally considered out of range for task-adequate responses (e.g. Greenwald, Nosek & Banaji, 2003). In addition, all trials with incorrect responses (7.9%) were discarded. In order to minimize the influence of individual differences in baseline response latencies in young children, an individual effect-size index of spontaneous gender stereotyping was used for the subsequent analyses. For this purpose, the individual's mean response latency in the congruent block was subtracted from the mean response latency in the incongruent block. The difference score was then divided by the pooled standard deviation of that individual's response latencies.

To test the influence of age on spontaneous gender stereotyping, the AIP scores were subjected to a 3 (Age Group) × 2 (Repetition) mixed-model ANOVA with Repetition as a within-subject factor contrasting the first and second parcels of 30 trials in each block.

Overall, spontaneous stereotyping scores differed significantly from zero, F (1, 61) = 15.74, $p < .01$. Children were faster in assigning toys in a stereotype-congruent than in a stereotype-incongruent manner. Note, however, that the order of stereotype-compatible and stereotype-incompatible blocks was not counterbalanced. Thus, scores of spontaneous gender stereotyping should be interpreted only in a relative manner (e.g. scores of one age group are different from scores of another age group), and not as reflecting absolute values of spontaneous gender stereotyping (e.g. scores higher than zero reflecting the absolute degree of spontaneous gender stereotyping). Although absolute latencies decreased from the first to the second half of each block, individual effect size scores remained stable, as indicated by a non-significant repetition effect, F (1, 61) = 2.11, $p = .15$, $\eta^2 = .03$. The critical test of a developmental change in spontaneous gender stereotyping was non-significant (see Table 1).

Combined analyses

To directly examine dissociations in the development of stereotype knowledge, stereotype flexibility, and spontaneous stereotyping across age groups, we tested the Age × Stereotype Measures interactions. To maximize comparability between stereotype measures, we limited the analyses to stereotype knowledge and stereotype flexibility scores based on toy pictures that were identical with the toy pictures in the AIP. Both stereotype knowledge and stereotype flexibility scores

Table 1 *Mean scores (and standard deviations) of gender stereotype knowledge, gender stereotype flexibility, and spontaneous gender stereotyping as a function of age*

	Age group			Age main effect	
	5 years	8 years	11 years	F (2, 59)	η^2
Gender stereotype knowledge					
Common objects (%)	78.5_A (11.3)	88.7_B (14.0)	96.5_C (5.7)	24.5***	.45
Toys (%)	98.5_A (3.3)	99.7_A (1.5)	99.7_A (1.0)	1.65 *ns*	.05
Gender stereotype flexibility					
Common objects (%)	33.8_A (23.1)	65.8_B (30.3)	87.9_C (16.6)	23.24***	.44
Toys (%)	14.5_A (22.8)	35.6_B (41.7)	77.8_C (34.3)	18.78***	.39
Spontaneous gender stereotyping					
AIP score	0.30_A (0.22)	$.45_A$ (0.28)	$.33_A$ (0.33)	1.66 *ns*	.05

Note: Row means with different subscripts (A, B and C) differ significantly (Student–Newman–Keuls, $p < .05$). Statistical tests of stereotype knowledge and stereotype flexibility were conducted using arcsin-transformed scores. * $p < .05$, ** $p < .01$, *** $p < .001$.

were first normalized using arcsin transformations. All scores (i.e. stereotype knowledge; stereotype flexibility; spontaneous stereotyping) were then *z*-transformed to obtain a common metric for all indices, and subjected to a 3 (Age Group) × 3 (Stereotype Measure) mixed-model ANOVA. As expected, the Age Group × Stereotype Measure interaction was significant, $F(4, 116) = 5.05$, $p = .001$, $\eta^2 = .15$. To further specify this interaction effect in terms of our hypotheses, we conducted follow-up analyses of the three pairs of stereotype measures. These analyses revealed that the obtained interaction was the result of the differential change of stereotype flexibility and spontaneous stereotyping, $F(2, 59) = 9.37$, $p < .001$, $\eta^2 = .24$, and of stereotype flexibility and stereotype knowledge $F(2, 59) = 4.42$, $p = .02$, $\eta^2 = .13$, whereas no differential change occurred between stereotype knowledge and spontaneous stereotyping ($F < 1$). As already revealed in our individual analyses, stereotype flexibility showed developmental change, whereas stereotype knowledge and spontaneous gender stereotyping remained constant across age groups.

Discussion

The present study shows a clear dissociation in the development of stereotype knowledge, stereotype flexibility and spontaneous stereotyping. Children exhibited a strong increase in gender stereotype flexibility from age 5 to 11 years. For toys and common objects, the proportions of the answer 'both' increased from the very low levels of 15% and 34% at age 5 to 78% and 88% at age 11. Stereotype knowledge, in contrast, already showed ceiling levels in the youngest age group, leaving little room for further increases as a function of age. These results replicate earlier meta-analytic results (Signorella *et al.*, 1992) and data from a longitudinal study (Trautner *et al.*, 2005). More interestingly, the stability of stereotype knowledge across age was paralleled by comparable stability of spontaneous stereotyping, which showed high levels irrespective of age. These findings are consistent with the assumption that spontaneous gender stereotyping is driven by stereotype knowledge, regardless of whether this knowledge is considered accurate or not (Devine, 1989; Strack & Deutsch, 2004).[2]

[2] One could argue that the proposed relation between stereotype knowledge and spontaneous stereotyping should be reflected in significant correlations between the two measures. Even though this reasoning is theoretically correct, it does not apply to the present data, which showed ceiling effects for stereotype knowledge scores across all age groups. Such ceiling effects significantly reduce variability, and thereby the possibility of detecting a factually existing relationship by means of correlation analyses (for a more detailed discussion of dissociation patterns, see Greenwald & Nosek, 2008).

The findings of the present study reveal an age-related dissociation of stereotype flexibility and spontaneous stereotyping, such that spontaneous stereotyping remained at a high level across age despite the fact that stereotype flexibility increased as a function of age. Verbal–behavioural dissociations of this kind are not uncommon in development (for a discussion, see Woolley, 2006) and provide a powerful means of gaining insight into the cognitive processes underlying overt behaviour. In the present case, evidence that age-related increases in stereotype flexibility were insufficient to reduce spontaneous stereotyping is consistent with dual-process models arguing that spontaneous stereotyping is influenced by the knowledge of the stereotype (Devine, 1989; Strack & Deutsch, 2004).

As such, the present findings point to the potential importance of automatic influences on overt social behaviour such as spontaneous stereotyping. Indeed, the basic logic of the AIP is that spontaneous action tendencies to respond in stereotype-congruent manner (e.g. to give dolls to girls and trucks to boys) facilitate quick responses when these tendencies are in line with the response required by the task. However, the same spontaneous action tendencies can interfere with quick responses when these tendencies are opposite to the one required in the task (i.e. give dolls to the boy and trucks to the girl). This response interference structure resembles those employed in many implicit measures used in social psychology, such as the Implicit Association Test (Greenwald, McGhee & Schwartz, 1998), affective priming (Fazio, Jackson, Dunton & Williams, 1995), or the extrinsic affective Simon task (De Houwer, 2003a). What all these measures have in common is that a prepotent response tendency may either facilitate or interfere with the response required in the task, which in turn influences the speed (and sometimes the accuracy) of participants' responses (De Houwer, 2003b; Gawronski, Deutsch, LeBel & Peters, 2008). The fact that even 11-year-old children, who exhibited high levels of stereotype flexibility, were slower to respond in a stereotype-incongruent manner than in a stereotype-congruent manner points to possible automatic influences on overt social behaviour.

It is important to keep in mind, however, that no behavioural task, including the AIP, is process-pure insofar as performance measures derived from such tasks necessarily reflect multiple underlying processes (for discussion, see Conrey, Sherman, Gawronski, Hugenberg & Groom, 2005). Thus, it remains possible that 5-, 8- and 11-year-old children showed comparable levels of spontaneous stereotyping in the AIP for different reasons. For instance, it is possible that the younger children may have had less knowledge about the gender stereotype than the older children, leading to a weaker tendency to respond in a stereotype-congruent manner compared with older children. At the same time, older children may have been better at inhibiting stereotype-congruent inclinations compared with

younger children (for a review, see Diamond, 2002). The net effect would have been comparable behavioural performance across older and younger children. Even though these speculations are inconsistent with the observed near-to-ceiling levels in stereotype knowledge across age groups, it would be useful to investigate the possibility of such dissociations by means of data analytical procedures that are capable of disentangling multiple underlying influences on behavioural outcomes (e.g. Conrey *et al*., 2005). Unfortunately, the AIP is not amenable to these data analytical procedures in its current form.[3] The development of modified variants of the AIP that meet the statistical requirements of these procedures would be an interesting avenue for future research.

In summary, the present results provide evidence for a dissociation in the developmental courses of stereotype flexibility and spontaneous stereotyping, suggesting that spontaneous stereotyping may be more closely related to stereotype knowledge. This finding is in line with the assumptions of contemporary dual-process models, which argue that stereotype knowledge can influence spontaneous behaviour even if this knowledge is considered inaccurate (e.g. Devine, 1989; Strack & Deutsch, 2004). In the present study, these assumptions were reflected in the similar stability at high levels of stereotype knowledge and spontaneous stereotyping across the age range from 5 to 11 years. This pattern deviated from the developmental course of stereotype flexibility, which showed a steady increase from age 5 to 11. To the degree that flexible use of stereotype knowledge implies a potential rejection of stereotype knowledge as inaccurate or wrong, the present findings highlight the insufficiency of stereotype flexibility for avoiding spontaneous stereotyping in children.

Acknowledgements

This research was partially supported by Academic Development Fund Grant 06-304 of the University of Western Ontario to Bertram Gawronski and J. Bruce Morton, and by Canada Research Chair Grant 202555 to Bertram Gawronski. The data of the Main Study are based on the MSc thesis of Hélène Gutt. Special thanks to Sarah Hart for her help in collecting the data for the Pilot Study.

[3] The quad-model is a multinomial model (see Batchelder & Riefer, 1999) that requires that the number of uniquely predicted response categories (e.g. a stereotypically female toy in the stereotype-congruent block) is higher than the number of process parameters estimated by the model, which is not the case for the AIP (for a more detailed discussion, see Conrey *et al*., 2005).

References

Bargh, J.A. (1994). The four horsemen of automaticity: awareness, intention, efficiency, and control in social cognition. In R.S. Wyer & T.K. Srull (Eds.), *Handbook of social cognition* (pp. 1–40). Hillsdale, NJ: Erlbaum.

Batchelder, W.H., & Riefer, D.M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, **6**, 57–86.

Bennett, M., & Sani, F. (2008). Children's subjective identification with social categories: a self-stereotyping approach. *Developmental Science*, **11**, 69–75.

Carter, D.B., & Patterson, C.J. (1982). Sex roles as social conventions: the development of children's conceptions of sex-role stereotypes. *Developmental Psychology*, **18**, 812–824.

Connor, J.M., & Serbin, L.A. (1977). Behaviorally based masculine- and feminine-activity-preference scales for preschoolers correlates with other classroom behaviors and cognitive tests. *Child Development*, **48**, 1411–1416.

Conrey, F.R., Sherman, J.W., Gawronski, B., Hugenberg, K., & Groom, C. (2005). Separating multiple processes in implicit social cognition: the Quad-Model of implicit task performance. *Journal of Personality and Social Psychology*, **89**, 469–487.

De Houwer, J. (2003a). The extrinsic affective Simon task. *Experimental Psychology*, **50**, 77–85.

De Houwer, J. (2003b). A structural analysis of indirect measures of attitudes. In J. Musch & K.C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 219–244). Mahwah, NJ: Lawrence Erlbaum.

Devine, P.G. (1989). Stereotypes and prejudice: their automatic and controlled components. *Journal of Personality and Social Psychology*, **56**, 5–18.

Diamond, A. (2002). Normal development of prefrontal cortex from birth to young adulthood: cognitive functions, anatomy, and biochemistry. In D.T. Stuss & R.T. Knight (Eds.), *Principles of frontal lobe function* (pp. 466–503). Oxford: Oxford University Press.

Diekman, A.B., & Eagly, A.H. (2000). Stereotypes as dynamic constructs: women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, **26**, 1171–1188.

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

Eisenberg, N., Murray, E., & Hite, T. (1982). Children's reasoning regarding sex-typed toy choice. *Child Development*, **53**, 81–86.

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

Fazio, R.H., Jackson, J.R., Dunton, B.C., & Williams, C.J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: a bona fide pipeline? *Journal of Personality and Social Psychology*, **69**, 1013–1027.

Gawronski, B., & Bodenhausen, G.V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. *Psychological Bulletin*, **132**, 692–731.

Gawronski, B., Deutsch, R., LeBel, E.P., & Peters, K.R. (2008). Response interference as a mechanism underlying implicit measures: some traps and gaps in the assessment of mental associations with experimental paradigms. *European Journal of Psychological Assessment*, **24**, 218–225.

Greenwald, A.G., McGhee, D.E., & Schwartz, J.K.L. (1998). Measuring individual differences in implicit cognition: the Implicit Association Test. *Journal of Personality and Social Psychology*, **74**, 1464–1480.

Greenwald, A.G., & Nosek, B.A. (2008). Attitudinal dissociation: what does it mean? In R.E. Petty, R.H. Fazio & P. Briñol (Eds.), *Attitudes: Insights from the new implicit measures* (pp. 65–82). New York: Psychology Press.

Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, **85**, 197–216.

Liss, M.B. (1981). Patterns of toy play: an analysis of sex differences. *Sex Roles*, **7**, 1143–1150.

Most, S.B., Sorber, A.V., & Cunningham, J.G. (2007). Auditory Stroop reveals implicit gender associations in adults and children. *Journal of Experimental Social Psychology*, **43**, 287–294.

O'Brien, M., & Huston, A.C. (1985). Development of sex-typed play behavior in toddlers. *Developmental Psychology*, **21**, 866–871.

Richardson, J.G., & Simpson, C.H. (1982). Children, gender, and the social structure: an analysis of the contents of letters to Santa Claus. *Child Development*, **53**, 429–436.

Serbin, L.A., & Sprafkin, C. (1986). The salience of gender and the process of sex typing in three- to seven-years old children. *Child Development*, **57**, 1188–1199.

Signorella, M.L., Bigler, R.S., & Liben, L.S. (1993). Developmental differences in children's gender schemata about others: a meta-analytic review. *Developmental Review*, **13**, 147–183.

Smith, E.R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, **4**, 108–131.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, **8**, 220–247.

Thompson, S.K. (1975). Gender labels and early sex role development. *Child Development*, **46**, 339–347.

Trautner, H.M., Ruble, D.N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: developmental or differential? *Infant and Child Development*, **14**, 365–381.

Widen, S.C., & Russell, J.A. (2002). Gender and preschoolers' perception of emotion. *Merrill–Palmer Quarterly*, **48**, 248–262.

Wilson, T.D., Lindsey, S., & Schooler, T.Y. (2000). A model of dual attitudes. *Psychological Review*, **107**, 101–126.

Woolley, J.D. (2006). Verbal–behavioral dissociations in development. *Child Development*, **77**, 1539–1553.

Received: 24 February 2008
Accepted: 20 January 2009