

Running head: AUTOMATIC PREJUDICE IN CHILDHOOD AND EARLY  
ADOLESCENCE

Automatic Prejudice in Childhood and Early Adolescence

Juliane Degner

University of Amsterdam

Dirk Wentura

Saarland University

Author note

Juliane Degner, Department of Social Psychology, University of Amsterdam, The Netherlands; Dirk Wentura, Department of Psychology, Saarland University, Germany. The research reported in this article was supported by a grant from the Deutsche Forschungsgemeinschaft (DFG) to Dirk Wentura and Juliane Degner (WE 2284/3-5). We thank Jasmin Schneider, Carmen Kühn, Anna Warnecke, and Jellie Sierksma for their assistance in data collection and Rahul Premraj and Skyler Hawk for thoughtful comments and stylistic corrections. Correspondence concerning this article should be addressed to Juliane Degner, Department of Social Psychology, University of Amsterdam, Roetersstraat 15, 1018WB Amsterdam, The Netherlands. E-mail: [jdegner@uva.nl](mailto:jdegner@uva.nl)

### Abstract

Four cross-sectional studies are presented that investigated the automatic activation of prejudice in children and adolescents (aged 9 to 15). Therefore, four different versions of the affective priming task were used, with pictures of ingroup and outgroup members being presented as prejudice-related prime stimuli. In all four studies, a pattern occurred that suggests a linear developmental increase of automatic prejudice with significant effects of outgroup negativity appearing only around the age of 12/13 years. Results of younger children, on the contrary, did not indicate any effect of automatic prejudice activation. In contrast, prejudice effects in an IAT showed high levels of prejudice independent of age (Study 3). Results of Study 4 suggest that these age differences are due to age-related differences in spontaneous categorization processes. Introducing a forced-categorization into the affective priming procedure produced a pattern of results equivalent to the one obtained with the IAT. These results suggest that although children are assumed to acquire prejudice at much younger ages, automatization of such attitudes might be related to developmental processes in early adolescence. We discuss possible theoretical implications of these results for a developmental theory of prejudice representation and automatization during childhood and adolescence.

(195 words)

Keywords: automatic attitudes, implicit prejudice, childhood, adolescence, development, affective priming, IAT

### Automatic Prejudice in Childhood and Early Adolescence

In the last two decades, the field of intergroup research has experienced a trend towards new social-cognitive theories and measures of prejudice and stereotypes. Considerable progress has been made in understanding the mental representation and cognitive processes underlying stereotypes and prejudice (for reviews see, e.g., Fiske, 2005; Mackie & Smith, 1998). For example, various studies have revealed how and under which conditions prejudice and stereotypes can be activated, without the perceiver's intent, merely by exposure to a relevant stimulus cue in the environment. It is a widely accepted assumption – yet little explored – that such automatic prejudice is rooted in early socialization experiences and learning processes in childhood: A broad and frequent use of stereotypes and prejudice from childhood onwards promotes their automatization (e.g., Devine, 1989). This assumption is in line with general theories of attitude automatization which is typically conceptualized as a slow-learning process based on gradual accumulation of trace experiences or well-learned associations (e.g., Fazio, Sanbonmatsu, Powell, & Kardes, 1986; Greenwald, & Banaji, 1995; Rudman, 2004). Interestingly, the underlying developmental processes of prejudice automatization have not yet been studied thoroughly (but see Baron & Banaji, 2006; Rutland, Cameron, Milne, and McGeorge, 2005 for notable exceptions), and the field of social cognition research lacks studies focusing on younger participants. The aim of the current research is to fill that void and explore automatic prejudice activation in children and adolescents. As the research was conducted in Germany and the Netherlands, we focus on attitudes towards Turkish and Moroccan immigrants as they constitute the

largest ethnic minority groups in these two countries (approx. 3.4 % of the German resident population; Woellert, Kröhnert, Sippel, & Klingholz, 2009; approx. 4.4% of the Dutch resident population; Centraal Bureau voor de Statistiek Nederlands, 2009). It has been repeatedly demonstrated that ten to twenty percent of German school children and one fourth to one third of Dutch and German adolescents openly express negative attitudes towards Turkish and Moroccan immigrants (e.g., Boehnke, Hagen, & Hefner, 1998; Frindte, Funke, & Waldzus, 1996; Masson & Verkuyten, 1993; Wagner, van Dick, Pettigrew, & Christ, 2003). If early acquisition and frequent use of prejudice throughout childhood and adolescence promotes their automatization (Devine, 1989), automatic prejudice should either come along with the open expression of prejudice or follow soon after. Nonetheless, no theory offers clear predictions about at which age, and under which conditions, automatic prejudice is to be expected.

#### Developmental theories of intergroup prejudice

Devine's dual process theory (1989) suggests that children of unspecified young age passively store parental and societal intergroup attitudes without being able to actively examine or dispute these attitudes. She assumes that these early acquired *societal* attitudes are stably represented in long term memory and form the basis on which later *personal* attitudes and beliefs are developed when the child is equipped with the necessary (yet unspecified) cognitive abilities. Because the early stored societal attitudes have a longer history of activation and are thus assumed to be represented as strong associative structures, they continue to influence the older child's and the adults thoughts, feelings and behaviors towards outgroups. Devine's theory concurs with

social developmental theories of intergroup prejudice (e.g., Aboud, 1988, 2008; Bigler & Liben, 2006; Nesdale, 1999, 2001) in that they all assume that social categories and their societal evaluations are acquired very early in age. Starting at the age of 3, children become able to distinguish and identify social categories, especially when these are perceptually distinguishable (e.g., by skin color), explicitly labeled (e.g., ‘Blacks, ‘African Americans’), salient by proportional group size (e.g., minority status) and explicitly and /or implicitly loaded with social significance (e.g., observed segregation and discrimination). This early awareness is typically accompanied by strong ingroup preferences that seem to increase until the age of 7/8 years. This early acquisition and use of societal evaluations of social groups and their members could thus be the basis of prejudice activation and automatization. Furthermore, Devine (1989) assumes that at a later age, dissociations between this (passively acquired) societal knowledge and more variable personal beliefs about social groups occur because of age-related expansion of cognitive ability and flexibility which in turn allows for the development of egalitarian and non-prejudiced views. Similarly, Aboud (1988, 2008) assumes in her social cognitive developmental theory that prejudice expression declines in older children (above 7 years) because of developmental shifts in children’s cognitive abilities which are related to their focus of attention (from the self to groups to individuals) and dominant information processes (from perceptual to affective to cognitive). Thus, the older child is assumed to be increasingly able to understand the individual rather than group-based characteristics of people and therefore develops more egalitarian views and reduces discriminative behaviors. According these theories, one would assume to find

automatic prejudice activation at a very early age with an increasing dissociation to deliberate prejudice expression above the age of seven.

Another social developmental theory, the Social Identity Development theory (Nesdale, 1999, 2001) challenges these assumptions. After reviewing existing research in intergroup biases in children, Nesdale points out that the early expressions of intergroup biases found in prejudice research are driven by mere ingroup preference (in forced choice research paradigms) and have little reference to negative evaluations as the basis of outgroup prejudice. He suggests that although well aware of societal prejudice at an early age (as suggested above), children adopt these outgroup prejudice as *personal attitudes* only later in life and with higher interindividual variability, related to (a) developmental processes of social identity formation and (b) children's varying perceptions of threat, tensions or conflicts between social groups. Only when awareness of own membership in a social group holding negative outgroup attitudes shifts into *active identification* with this ingroup, children approve these attitudes as their own depending on their varying perceptions of intergroup relations. Nesdale assumes the starting age for the shift in social identity and intergroup attitudes around the age of 7 years. This is in line with research associating late childhood and early adolescence as a critical phase of active identity development which includes increasing exploration and formation of one's own social (ethnic) identities (e.g., French, Seidman, Allen, & Aber, 2006; Newman & Newman, 2002; Phinney & Chavira, 1992). During these years, children and adolescents deliberately face their racial, ethnic, cultural, national, and sexual identities, and begin to more consciously identify with their ingroups. This

identification is accompanied by an internalization of ingroup values, beliefs, and attitudes, as well as acknowledgment and accentuation of differences between own and other groups. One could thus assume that social categories become *personally and socially meaningful* only during late childhood and early adolescence and are from then on more frequently used in relation to social identity development (Nesdale, 2001). If prejudice automatization relies on frequent activations and use of outgroup categorization and evaluation, the social identity development perspective would predict a far later onset of automatic prejudice with a much higher interindividual variability. Thus, from current theorizing, opposing predictions about the onset of automatic prejudice can be derived; with an early onset following passive acquisition and non-deliberate use of societal intergroup attitudes in early childhood vs. a late onset following active acquisition and motivated use in late childhood and early adolescence.

Hitherto, there are only few studies that explored automatic prejudice in younger participants. Using a child-friendly version of the Implicit Association Test (IAT, Greenwald, McGhee, & Schwartz, 1998) in a cross-sectional study, Baron and Banaji (2006) found that six-year-old children showed the same amplitude of automatic prejudice as 10-year-olds and young adults aged 19 years. Similarly, Rutland and colleagues (2005) found evidence of implicit intergroup prejudice using an IAT with children aged from 6 to 8, and 10 to 12, as well as adolescents aged 14 to 16 years. Not only were IAT scores of the same size across age groups, but they were also unaffected by children's knowledge of social norms or concerns for self-presentation. The finding of automatic prejudice at such early age clearly supports the first perspective: It might

be concluded that automatic prejudice is acquired soon after the age at which children learn the conception of ethnic categories, and remains relatively stable across development, stored in memory as stable valence associations. Automatization of negative intergroup attitudes would thus not depend on frequent recall or expression of negativity toward outgroups and their members (e.g., Banaji, Baron, Dunham, & Olson, 2008).

#### Assessing automatic prejudice

It is noticeable that the above cited studies relied on the IAT, currently the most prominent response time-based indirect measure of attitudes in social psychology. The IAT involves repeatedly categorizing target stimuli according their social category membership (e.g., Black vs. White) and attribute stimuli according their valence (i.e., positive vs. negative), whereas the response-category assignments are varied (see Greenwald et al., 1998, for a detailed description). Without discrediting the valuable results found using the IAT, it remains a worthwhile endeavor to expand our knowledge about developmental gradients of automatic prejudice-related evaluations by using other measures. Related to the aforementioned theoretical considerations, there are several reasons why IAT results should be complemented by research using other measures of automatic prejudice. First, the IAT is a comparative measure, and does not allow distinction of ingroup positivity from outgroup negativity. The documented stable IAT effect across age groups might be the result of a stable ingroup positivity, of a stable outgroup negativity, or of complementary developmental changes in ingroup and outgroup evaluations. For example, if decreasing ingroup positivity was accompanied

by increasing outgroup negativity, a stable IAT effect would be observed but essential developmental changes would remain hidden. Second, the IAT is a category-based measure, in that it assesses evaluative reactions to given category labels (e.g., Blacks and Whites). However, negative evaluations of a social category do not necessarily correspond directly to evaluative reactions toward individual exemplars of that category (see Livingston & Brewer, 2002; Olson & Fazio, 2003). It is conceivable that children's aforementioned early awareness of societal evaluations of social categories translate into corresponding category-based IAT scores when group labels are made salient, which may or may not be related to automatic negative reactions toward *individual exemplars* of that category. Also, categorizations in the IAT are forced-choice: The target has to be categorized into one of two given groups. Without further research we do not know for sure that these categories are actually the ones children spontaneously use as a basis of impression formation in daily life encounters. Third, there is an ongoing debate about the extent to which IAT effects are related to factors other than associations between concepts (see De Houwer, Teige-Mocigemba, Spruyt & Moors, 2009, for a recent review). For example, it has been proposed that effects of stimulus salience, familiarity, and perceptual fluency might play a substantial role in producing effects in the IAT (e.g., Brendl, Markman, & Messner, 2001; Chang & Mitchell, 2009; Rothermund & Wentura, 2004). From this critical perspective, stable IAT effects across different age groups would demonstrate a stable difference between ingroup and outgroup members regarding their salience, familiarity, and/or perceptual fluency, and do not necessarily point to evaluative associations. Finally, there is some evidence that

the IAT might be susceptible to demand effects, such as when adult participants produce desired IAT scores in relation to novel objects if they think the experimenter expects them to do so (De Houwer, Beckers, & Moors, 2007). This has important implications for the usage of the IAT for exploring developmental aspects of attitudes. For example, it cannot be ruled out that children assume that they are asked to show negative evaluations of outgroup members if that seems to be the societal standard. Thus, it cannot be distinguished whether a given score in an IAT reflects a child's personal prejudice or her or his awareness of societal prejudice (see Han, Olson, & Fazio, 2006; Olson & Fazio, 2004; for a related argument).

The aforementioned issues illustrate that complementing research using other measures is needed in order to get a clearer picture of prejudice acquisition and automatization throughout childhood and adolescence. For our studies, we decided to use another prominent indirect attitude measure, the affective priming task (Fazio et al., 1986; Fazio, Jackson, Dunton, & Williams, 1995). In the affective priming task, two stimuli are presented in short succession, the first being the prime, the second being the target. The main task of the participants is to categorize the targets with regard to their valence as positive or negative. Although the valence of the prime is completely irrelevant for the task, it is known that it promotes the response to a target of same valence whereas it interferes with the response to a target of opposite valence. Thus, differences in response latencies to positive versus negative targets in relation to attitudinal primes allow inferring the prime valence. The underlying mechanism of affective priming effects seem to be rather well explained, and directly linked to

automatic activation of prime valence (De Houwer et al., 2009; Klauer & Musch, 2003; Wentura & Degner, in press). The affective priming task offers a number of virtues for the current research question. First, it is assumed that affective priming effects are primarily determined by the attitudes towards the individual prime stimuli, because participants do not have to construe primes in terms of their social category memberships. Thus, the paradigm permits conclusions about the degree to which exemplar outgroup members automatically activate evaluations when they are presented as primes (Fazio & Olson, 2003; Livingston & Brewer, 2002; also see De Houwer, 2001). Nonetheless, cover story and task instructions can be varied in a way that instigate category-based vs. exemplar-based processing of primes (Fazio & Olson, 2003). Second, the design of the affective priming task allows separate consideration of age differences in ingroup vs. outgroup evaluations, although this should be done with caution (see Scherer & Lambert, 2009; Wentura & Degner, in press). Third, the procedure allows primes to be presented outside of participants' conscious perception, and demand effects or any effects of intentional modulation can therefore be avoided (at least with masked prime presentation; see Degner, 2009). Also, it has been shown that affective priming effects are less susceptible to normative information or prevalent societal attitudes (Han et al., 2006). Thus, priming effects might be straightforwardly interpreted as indicators of personal attitudes instead of mere awareness of societal attitudes. In conclusion, even though affective priming and the IAT both belong to the class of indirect response latency-based ("implicit") attitude measures, they do not necessarily tap the same underlying processes and/or concepts. Although finding

converging results will certainly be considered as mutual evidence for validity, diverging results would suggest a more differentiated pattern of processes related to automatic social evaluations in children and adolescents.

### Overview

To complement existing studies on age effects in automatic prejudice, we used the affective priming task in four cross-sectional studies with German and Dutch school children, aged from 9 to 15 years (overall sample size  $N = 833$ ). As the priming measure is based on a simple evaluative categorization task, we expected it to be easily accomplished by children of these age groups. To our knowledge, no studies have yet used this paradigm with participants of younger age. Thus, the general applicability of the measure to younger participants had to be demonstrated. We therefore included a standard priming condition using normatively positive and negative primes, which should result in the typical affective priming effects. For the assessment of prejudice-related evaluations, we included pictures of persons of Turkish (Study 1 to 3) and Moroccan (Study 4) origin as outgroup primes and pictures white Germans (Study 1 to 3) and Dutch people (Study 4) as ingroup primes.

The main focus of the studies was to explore whether automatic prejudice activation in the affective priming task differs according to the age of participants. Study 1 and Study 2 complement one another: In Study 1, we used briefly presented, but clearly perceptible, pictures of Turks and Germans as primes. In Study 2, however, we used masked presentation of primes thus precluding conscious identification of the primes. In Study 3, the aim was to make a direct comparison between the affective

priming task and the IAT. In Study 4, we tested the assumption that age differences found in automatic prejudice activation in the affective priming task are related to children's use of ethnic categories in person perception.

### Study 1

In Study 1, we employed an affective priming task with portraits of young men of German vs. Turkish origin presented as clearly visible primes. To establish whether the affective priming paradigm is generally applicable for the assessment of automatic evaluations in children, we implemented a control condition using standard positive and negative stimuli as primes (i.e., emotional pictures). Thus, the main goals of Study 1 were (a) to ascertain the general applicability of the affective priming paradigm with children and (b) to explore age differences of automatic prejudice-related evaluations in the paradigm. In addition, we tried to implement a between participants manipulation of prime construal via different instructions in a prime memorization task, intended to initiate category-based vs. individual-based processing of prime pictures (see Olson & Fazio, 2003). However, already during data collection, we observed that children ignored or forgot the memorization task, which eventually proved to have no influence on priming effects. We will therefore not discuss this manipulation in detail. However, we return to this issue with Study 4, where we implemented a successful manipulation of prime construal.

### *Method*

#### *Participants*

Participants were 304 students from grades five to eight in two public German high schools. The schools were situated in two mid-sized towns in South-Western Germany with comparable percentages of immigration in the city populations (12.9% and 9.6%, respectively). Approximately 3 to 5% of the school populations were from a visible ethnic minority, and all teachers were white German. The sample is divided into 81 fifth-graders (46 boys, 35 girls; median age = 11, ranging from 10 to 12), 84 sixth-graders (45 boys, 39 girls; median age = 12, ranging from 11 to 13), 91 seventh-graders (64 boys, 27 girls; median age = 13, ranging from 12 to 14), and 48 eighth-graders (30 boys, 18 girls; median age = 14, ranging from 12 to 15). Prior to the study, we obtained written parental consent for participation. As we focus on development of intergroup bias in majority children, the data of 45 further students with various immigration backgrounds (German was not their native language and/or they were not born in Germany) were discarded from analyses. Data of another 7 participants had to be excluded because of individual circumstances impeding data collection (i.e., arm fraction, visual impairment, attention deficits). Each class received a reward of 50 EUR for communal use, regardless of the number of participating children. The study was conducted in the computer classroom of the schools during the last three weeks of the school year prior to the summer vacation.

### *Materials*

*Priming measure.* The priming task conformed to a 2 x 5 within-subjects design, made up by the factorial combination of target valence (positive vs. negative) and prime type (positive vs. negative vs. neutral vs. Turkish vs. German). For the prejudice-related

primes, portrait pictures of 10 Turkish and 10 German young men were used. These pictures were selected from a large pool of 200 portrait pictures, after intensive pre-testing with university students (see Degner Wentura, Gniewosz, & Noack, 2007, for a description of pretests). A further pretest was conducted in which 53 children from two fifth- ( $n = 27$ ) and seventh-grade classes ( $n = 25$ ) had to rapidly categorize the depicted men as Turkish or German (Cronbach's  $\alpha = .83^1$ ). Only pictures with fast responses ( $\leq 1000$  ms) and a minimum of 90% correct responses were included as primes in the current experiment. Via these additional pretests, it was assured that the depicted individuals could be easily and unambiguously categorized as being Turkish and German on first sight, also by the younger participants in this study. The Turkish and German prime sets were matched according their intra-group typicality, their attractiveness, and the emotionality of their facial expressions in order to avoid the possibility that variables other than ethnicity could account for differential priming effects. All faces were in frontal view against a white background. They were in black and white, and approx. 75 mm high and 50 mm wide. To obtain a reference priming effect with standardized stimuli, pictures from the International Affective Picture System (IAPS, Lang, Bradley, & Cuthbert, 2005) were included as additional primes; thereof five with positive, negative, and neutral contents, respectively. Positive primes were characterized by a mean valence of  $M = 7.86$  ( $SD = .31$ ) and a mean arousal value of  $M = 4.95$  ( $SD = .23$ ). Negative primes had a mean valence of  $M = 2.91$  ( $SD = .53$ ) and a mean arousal value of  $M = 5.60$  ( $SD = 1.03$ ). Neutral primes had a mean valence of  $M = 5.02$  ( $SD = .11$ ) and a mean arousal value of  $M = 2.76$  ( $SD = .15$ ).

The target set for the priming task consisted of 50 IAPS pictures, 25 with a positive content and 25 with a negative content. Positive targets were characterized by a mean valence of  $M = 7.64$  ( $SD = .35$ ) and a mean arousal value of  $M = 4.97$  ( $SD = 1.04$ ). Negative targets had a mean valence of  $M = 3.09$  ( $SD = .74$ ) and a mean arousal value of  $M = 5.24$  ( $SD = .86$ ). To assure that positive and negative pictures chosen from the IAPS were distinctly and unanimously positive or negative for children of the respective age groups, a speeded valence classification task had been included in the aforementioned pretest with children. Only pictures with high agreement ( $> 90\%$ ) in this task were chosen as stimuli for the current study.

*Explicit measures.* To assess participants' open expression of prejudices toward Turks living in Germany, items of the Blatant Prejudice Scale (Pettigrew & Meertens, 1995) were reformulated for children (e.g., "*I would not mind if a Turkish person was appointed as my teacher as long as he is giving good classes.*", see Degner et al., 2007). A scale to assess general xenophobic attitudes (e.g., "*Foreigners take away our jobs*", see Balke, El-Menouar, & Rastetter, 2002) was also included. All items were answered on a 6-point Likert scale ranging from -3 ("I do not agree at all") to +3 ("I totally agree"). Internal consistencies of both questionnaire scales were in the range of expectations, with Cronbachs'  $\alpha = .74$  and  $.78$  for the blatant prejudice scale (in Study 1 and 2) and  $\alpha = .83$  and  $\alpha = .74$  for the anti-foreigner scale (in Study 1 and 2, respectively). Interscale correlation was  $r = .64$ ,  $p < .001$  in Study 1 and  $r = .24$ ,  $p < .001$  in Study 2, indicating that the two scales measured only partially overlapping concepts.

### *Procedure*

Participants were tested in groups of 6 to 14 students seated separately in a computer classroom of their schools. They were informed that they would accomplish a series of tasks on the computer and fill out a questionnaire. All instructions were given by one of two female experimenters using presentation slides for illustration.

The affective priming task was introduced as a mixed test of evaluation and memory abilities. Children were introduced to the complete task in three steps. In the first part, children were familiarized with the target evaluation task. They were informed that pictures would be presented very briefly on the computer screen and their task was to quickly categorize the pictures according their valence by means of a key press (with [A] = negative and [5] on the number pad = positive). Participants completed two practice blocks of 25 trials each. Each trial started with a fixation cross presented for 317 ms on the screen. After a blank of 133 ms, the target pictures appeared and remained on screen until participants responded. To encourage fast responding, the picture remained on the screen maximally for 1750 ms, after which it was replaced by a feedback "*zu langsam*" (too slow). Participants received instantaneous accuracy feedback after each trial. Furthermore, participants received a summarized feedback on their percentage of correct responses and mean response time at the end of the practice phase. The experimenters explained the feedback and encouraged participants to ask any questions about the task.

In the second step, participants were familiarized with the memory task. They watched a sequence of 10 pictures of young Turkish and German men, interspersed with

15 IAPS pictures with positive, negative, and neutral contents. Participants were instructed to memorize the person pictures. Participants received different instructions for the memory task, depending on their assigned experimental condition. Participants within a particular session received the same instructions in order to ensure their unawareness of the experiment's purpose. Participants' assignments to the sessions were randomly defined. In the *Categorization* condition, children were instructed to memorize the number of Turks and Germans for a later test. In the *Personalization* condition, children were instructed to memorize the individual faces. A memory test followed directly after the memorization phase. Participants in the *Categorization* condition reported the number of Turks and Germans they remembered seeing. Participants in the *Personalization* condition saw the earlier person pictures interspersed with new pictures, and pressed an "old" key when they had seen the picture before and a "new" key when the picture was novel.

In the third, experimental phase of the affective priming task, participants were informed that the evaluation task and the memory task they had practiced so far would now be combined. They would therefore always be presented with two pictures in short succession: the first picture (if it was a person picture) being target for the memory task, the second being target of the evaluation task. The experimenters repeated the instructions for both tasks (see above) and proceeded only when they were sure that all children had fully understood all instructions. The experimental phase consisted of four blocks, each containing 50 trials. Each trial started with the presentation of the prime stimulus for 317 ms, replaced by a blank screen. The target stimulus followed after 133

ms (Stimulus Onset Asynchrony [SOA] = 450 ms) and remained until the participants responded or the response deadline had elapsed.<sup>2</sup> The inter-trial interval was 1000ms. Within a block, each prime was presented once in each target condition. During the whole experiment, participants received instantaneous accuracy feedback and summarized feedback at the end of each block (i.e., mean response times and percentage of correct responses). The priming task took approximately 20 min. After a small break and a filler task (app. 10 minutes), the experimenter handed out the questionnaires and explained the item format and answering scales with two neutral exemplar items (e.g. "*I prefer lemonade over water.*"). All items of the questionnaire were then read aloud by the experimenter, and participants filled in their responses on individual sheets. Participants then provided demographic data, followed by a full debriefing.

### *Results*

Analyses will be presented separately for questionnaire and priming data.

#### *Explicit measures*

To analyze age effects, scale means were included in a 2 (Scale: blatant prejudice vs. xenophobia) by 4 (Grade: 5. vs. 6. vs. 7. vs. 8.) analysis of variance with repeated measures on the first factor. A significant main effect of scale was found,  $F(1,300) = 54.67, p < .001, \eta^2 = .154$ , the main effect of grade missed the conventional level of significance,  $F(3,300) = 2.47, p = .06, \eta^2 = .024$ . Main effects were qualified by a significant interaction,  $F(3,300) = 4.24, p = .006, \eta^2 = .041$ . General xenophobic attitudes were somewhat higher than blatant prejudice against Turks, and did not differ according to age groups,  $F(3,300) = 1.42, ns$  (see Table 1). On the contrary, blatant

prejudice against Turks depended on participants' age,  $F(3,300) = 4.03, p = .008, \eta^2 = .039$  (linear trend  $F(1,300) = 9.71, p = .002$ )<sup>3</sup>. Older participants expressed lower levels of explicit prejudice, as compared to younger participants. Participants' gender did not affect prejudice expression,  $F < 1, ns$ .

### *Priming effects*

Response latency data were trimmed to correct for anticipatory responses and momentary inattention. Trials with outlying response latencies slower than 1.5 interquartile ranges above the third quartile of the individual RT distribution (see Tukey, 1977) or below 300 ms were excluded from analyses (6.8%). Analyses were run on response latencies of trials with correct responses only. The mean error rate was 9.54% ( $SD = 8.01$ ); no systematic influences of prime presentation on error rates were found,  $F_s \leq 1.40, p_s \geq .24$ .

Priming indices were calculated by subtracting the mean response latencies in congruent trials from incongruent trials, thus representing the crucial interaction of prime valence and target evaluation. Accordingly, priming indices for the standard primes represent the difference of mean RTs for trials with the same valence of prime and target (i.e., positive-positive, negative-negative), as compared to mean RTs of trials with primes and targets of different valence (positive-negative, negative-positive). The expected response facilitation in congruent trials, as compared to incongruent trials, would thus be indicated by a positive priming score. Likewise, priming indices were computed for the attitude-relevant primes, taking the German/positive and Turkish/negative pairings as congruent combinations. Thus, positive priming scores

represent higher prejudice in terms of relative negativity towards Turks compared to Germans. Priming effects for standard and attitude-related primes were analyzed separately.

In the standard priming condition (i.e., IAPS pictures as primes), we found a significant priming effect of  $M = 13$  ms,  $SD = 36$ ,  $t(303) = 6.21$ ,  $p < .001$ ,  $d = 0.36$ , indicating that participants reacted faster in trials with a valence-congruent prime-target combination as compared to incongruent pairings. Age did not affect the priming score,  $F(3, 300) = 0.56$ , *ns*; significant priming effects of comparable size were found in all age groups (all  $ts \geq 2.30$ ,  $ps \leq .03$ ,  $0.31 \leq d \leq 0.43$ ). The standard priming effect was neither affected by participants' gender, nor by type of memory task, both  $Fs < 1$ , *ns*. This pattern of results indicates that the priming paradigm yielded the expected effects: prime valence systematically influenced target evaluations, regardless of age or sex of participants and regardless of memory task. It can therefore be concluded that the priming task was sensitive for effects of automatic valence activation in this sample of children and adolescents.

For the prejudice-related prime pictures, a significant effect of grade occurred,  $F(3,300) = 3.02$ ,  $p = .03$ ,  $\eta^2 = .03$  ( $F(1,300) = 5.47$ ,  $p = .02$ , for the linear trend). As can be seen in Figure 1, participants in the younger grades did not show significant priming effects ( $M_{5th} = -2$  ms,  $SD = 36$ ,  $t(80) = -0.48$ ,  $p = .63$ ;  $M_{6th} = 4$  ms,  $SD = 25$ ,  $t(83) = 1.61$ ,  $p = .11$ ), whereas participants in the higher grades displayed significant priming effects ( $M_{7th} = 9$  ms,  $SD = 30$ ,  $t(90) = 2.93$ ,  $p = .004$ ,  $d = 0.30$ ;  $M_{8th} = 9$  ms,  $SD = 20$ ,  $t(47) = 3.32$ ,  $p = .002$ ,  $d = 0.48$ ). Thus, whereas the reactions of the younger children were not

differentially affected by the presentation of Turkish vs. German primes, a systematic priming effect could be found for the adolescents. Planned contrasts (Helmert) were calculated to test for the occurrence of significant differences of priming effects between age groups. Therefore, the priming score of each age group (starting with the 8<sup>th</sup> grade) was tested for difference from priming scores of the younger age groups, respectively. This contrast was significant for the 7<sup>th</sup> grade,  $F(1,300) = 4.40, p = .04$ , but not for the 6<sup>th</sup> and 8<sup>th</sup> grades,  $F_s < 1.95, p_s > .16$ . These results indicate a developmental shift between 6<sup>th</sup> and 7<sup>th</sup> grade. The prejudice-related priming effect was neither affected by participants' gender, nor by type of memory task, both  $F_s \leq 1.36, ns$ .

To explore the extent to which the age effect upon prejudice-related priming resulted from automatic in-group preference vs. out-group derogation, we computed separate priming indices for Turkish and German primes. Therefore, mean response latencies in trials with positive targets were subtracted from trials with negative targets. Note that the *absolute* value of these indices can not be straightforwardly interpreted as direct indices of positive vs. negative attitudes, because they are distorted by main effects of target valence (e.g., reactions to negative targets are typically slower as compared to positive targets, see, e.g., Wentura & Degner, in press). Still, these difference scores can be used to explore differences in age trends for Turkish and German primes.<sup>4</sup> The priming index for German primes was unrelated to participants' age,  $F < 1, ns$ , indicating that automatic evaluations of German pictures were similar for participants of all age groups ( $M = 29$  ms,  $SD = 50$ ). On the contrary, the priming indices for Turkish pictures were significantly related to age,  $F(3, 300) = 4.26, p = .006$ ,

$\eta^2 = .041$ , indicating that evaluative reactions to Turkish primes caused the age effect in priming indices ( $M_{5th} = 36$  ms,  $SD = 64$ ;  $M_{6th} = 18$  ms,  $SD = 39$ ;  $M_{7th} = 10$  ms,  $SD = 54$ ;  $M_{8th} = 12$  ms,  $SD = 39$ ).

As an estimate of internal consistency of the priming measure, we calculated split-half partial correlations for the Turkish priming index controlled for the German priming index for 30 independent random splits of the trial sequence. The median Spearman-Brown-adjusted reliability score was  $r = .31$ , replicating findings on low internal consistency of the affective priming paradigm (e.g., Bosson, Swann & Pennebaker, 2000; De Houwer, 2003; but see Cunningham, Preacher & Banaji, 2001). There were no significant correlations of composite or separate priming scores with self-report measures, neither for blatant prejudice nor for general xenophobia,  $-.01 \leq r_s \leq .06$ , *ns*.

#### *Memory performance*

As already noted in the introduction to this study and as evident from the null results concerning this factor, our manipulation of prime categorization versus personalization was supposedly too weak to influence priming processes in Children. Nevertheless, we briefly summarize basic results of the memory tests for the sake of completeness.

The memory test for participants in the *Categorization* condition consisted of an estimation of the number of Turkish and German pictures that they had seen during the priming task. Overall, participants tended to slightly underestimate the percentage of Turkish primes in comparison to German primes,  $M = 46.23\%$   $SD = 9.22$ , with  $t(139) = -4.85$ ,  $p < .001$ , for the deviance from 50%. Category estimates were not affected by

participants' age,  $F(3,136) = 1.85, p = .14$ , and were unrelated to affective priming effects,  $r = .02, ns$ .

The memory test for the *Personalization* condition asked for participants' judgments of whether or not various portraits of Turks and Germans had been presented during the previous priming task. As an indicator of memory performance, the non-parametric signal detection sensitivity index  $A'$  (Pollack, 1970) was computed for Turkish and German primes, separately. Values were tested for a positive deviation from .50, indicating better-than-chance performance in recognizing the prime pictures. Memory performance was not significantly different for German and Turkish primes,  $F(1,160) = 2.38, ns$ . Not surprisingly, memory performance was significantly related to participants' age,  $F(3,160) = 7.00, p < .001$ , but no interaction between prime-type (Turkish vs. German) and age occurred,  $F(3,160) = 1.03, ns$ . Memory performance of fifth-graders did not differ from chance, with  $M = .53, SD = .12, t(42) = 1.66, p = .11$ . Sixth- and seventh-graders performed slightly better than chance  $M_{6th} = .56, SD = .14, t(43) = 2.61, p = .01$  and  $M_{7th} = .57, SD = .15, t(46) = 3.19, p = .003$ . Eighth-graders' performance clearly exceeded the chance level,  $M = .67, SD = .11, t(29) = 8.32, p < .001$ . Again, memory performance scores were not related to affective priming effects,  $r = .09, ns$  and  $r = .04, ns$ , (and  $r = .04$  for their difference score), and thus did not mediate the age effects on priming scores.

### *Discussion*

In this study, we showed that an affective priming task is generally applicable as an indirect measure of children's and adolescents' automatic activations of evaluations. To

the best of our knowledge, this is the first time that an affective priming effect was shown in participants younger than the typical college student samples. The significant standard priming effect found with normative positive and negative primes demonstrates that the technique is sensitive to effects of automatic valence activation in the studied age groups. Independent of participants' age, responses to valenced stimuli were facilitated or inhibited as a function of the valence congruence or incongruence of the preceding primes.

The implication of this standard effect is that when pictures of Germans and Turks are used as primes, their facilitative or inhibitory effects on the target categorization allow for an estimation of automatically activated evaluations of ingroup and outgroup members. This prejudice-related priming effect was significantly related to participants' age. Younger participants – attending fifth and sixth grade – showed no differences in spontaneous evaluations of Turkish and German prime pictures. Older participants – attending seventh and eighth grade – yielded significant priming effects indicating the presence of automatically activated evaluations and prejudice.

The age effects found for automatic prejudice activation stand in sharp contrast to open prejudice expression. The overall level of open prejudice was rather low, and expression of negativity tended to decrease among participants as they increased in age. It thus seems that developmental characteristics differ for automatic prejudice activation and explicit prejudice expression.

Interestingly, priming effects were not influenced by the manipulation of prime construal, as introduced by the different memory tasks. The results of the memory test

reveal that overall memory performance was generally rather poor. A very plausible explanation (supported by the experimenters' observations during data collection) is that children abandoned the memory task while focusing on the target evaluations, because they received immediate accuracy feedback only on the latter. This would also explain why we found no effect of categorization vs. personalization in the priming task. We will return to this issue with Study 4. Still, even if widely ignored, the memory tasks might have partially shifted participants' attention toward the primes. One might even argue that the attention spent on primes can somehow account for the found age differences in the priming effects: Whereas young participants might naively not have considered prejudice in this task, older participants might have become aware of the prejudice-relatedness of the primes, which prompted them to actively activate evaluations of the primes. Although we believe this to be rather unlikely, it seems to be worth exploring whether priming effects – and the respective age differences – also occur in children when they devote no attention to the primes. Thus, to complement the results of Study 1, we conducted a second study using the affective priming task with masked prime presentation and a shorter SOA. Masking primes excludes participants' conscious awareness of the priming event (i.e., subliminal priming); thus, demand effects or any sort of intentional activation of prejudiced evaluations can be precluded. Finding priming effects under such conditions would allow for the straightforward interpretation that evaluations being indirectly measured with the affective priming are truly automatic, in the sense of being unintentional and unconscious. Although we found masked effects of automatic evaluation in former studies with adults and

adolescents (e.g., Degner et al., 2007; Degner & Wentura, 2009; Otten & Wentura, 1999; Wentura, Kulfanek, & Greve, 2005), it is important to replicate masked standard affective priming effects with children.

### Study 2

In Study 2, we conducted a modified version of the affective priming task, using masked prime presentation. The experiment was conducted with children of the same age groups as in Study 1.

### *Method*

#### *Participants*

Participants were 264 students from grades five to eight from a different public high school as in Study 1 (situated in the same region). Participants were comprised of 68 fifth-graders (34 boys, 34 girls; median age = 10, ranging from 9 to 11), 53 sixth-graders (26 boys, 27 girls; median age = 11, ranging from 11 to 12), 82 seventh-graders (34 boys, 48 girls; median age = 12, ranging from 11 to 13), and 61 eighth-graders (25 boys, 36 girls; median age = 13, ranging from 12 to 14). The data of 13 additional participants with immigration backgrounds were discarded from analyses. Data of seven further participants were incomplete due to computer failure, and data of another five participants had to be excluded because of individual circumstances that impeded data collection. The study was conducted at the beginning of the school year, which explains why students of the same grades were approximately one year younger in Study 2, as compared to Study 1.

#### *Materials and Procedure*

We used the same questionnaires and priming stimuli as in Study 1. The procedure was similar to Study 1, with the following exceptions: The affective priming task was introduced as a test of evaluation and categorization abilities. That is, there was no memorization task and the existence of primes was not revealed to participants. Prime presentation was reduced to 30 ms. To assure participants' unawareness of prime presentation, the prime was embedded into a forward mask, presented for 70 ms, and a backwards mask, presented for 10 ms; thus the SOA amounted to 40 ms. The subjective impression of this presentation sequence was a brief flicker. Since it is known from masked affective priming studies that effects are more robust if participants are forced to respond faster as needed to maintain the typically high accuracy (see, e.g., Draine & Greenwald, 1998), we employed a response-deadline technique. That is, the target remained on screen for only 500 ms and children were instructed to press the correct key as fast as possible within the span of target presentation. If they did not respond within this time span, the target was replaced by a blue sign, which signaled that the response was too slow. Such modifications largely remove any variance of interest from latency data. Therefore, in accordance with other studies (e.g., Degner et al., 2007; Draine & Greenwald 1998), the error rate was the dependent variable of interest in Study 2.

After the priming task, additional ratings of liking were obtained for all Turkish and German prime pictures. To acquire these ratings, the pictures were presented on screen in a random order and participants indicated their liking for each portrayed person on a Likert scale ranging from 1 (*do not like at all*) to 7 (*like very much*). Children were then

interviewed about prime awareness in a funneled debriefing procedure. Therefore, children were first informed about the hidden presentation of primes, and then asked to indicate whether they had noticed the priming during the priming task (yes vs. no), whether they had perceived the hidden pictures (yes vs. no), and if yes, whether they could list the content of the pictures they had recognized. Subsequently, a direct test of prime recognition was administered. Participants worked through 30 priming trials with the standard positive and negative primes and 30 priming trials with the Turkish and German primes. They were instructed to identify the prime pictures and to categorize them as being either positive vs. negative or Turkish vs. German. To assure comparable masking without showing the valence-connoted targets as during the priming task, a neutral IAPS picture was presented as target stimulus in each test trial. Thereafter, questionnaires were completed after a filler task as in Study 1.

### *Results*

#### *Explicit measures*

*Questionnaires.* No significant effects of age on open prejudice expression in the questionnaires were found,  $F$ 's < 1.18, *ns.* (see Table 1). Contrary to Study 1, students of all grades expressed similar low levels of blatant prejudice towards Turks,  $M = 3.01$ ,  $SD = 0.52$ , and general xenophobic attitudes,  $M = 3.12$ ,  $SD = 0.96$ .

*Stimulus Ratings.* The pleasantness ratings of the portrait pictures were submitted to a 2 (Prime: Turkish vs. German) by 4 (Age) analysis of variance. We found a significant main effect of prime origin on the ratings,  $F(1,260) = 212.39$ ,  $p < .001$ ,  $\eta^2 = .452$ , indicating that German stimuli ( $M = 4.73$ ,  $SD = .88$ ) were rated significantly higher in

pleasantness than Turkish stimuli ( $M = 3.59$ ,  $SD = 1.14$ ). This main effect was qualified by a marginally significant interaction effect with age,  $F(3,260) = 2.17$ ,  $p = .09$ ,  $\eta^2 = .024$ . Whereas ratings of German stimuli were equally positive in all age groups,  $F < 1$ ,  $ns$ , ratings of Turkish stimuli were positively related to age, such that older participants rated them as being less negative than did younger participants,  $F(3,260) = 2.61$ ,  $p = .05$ ,  $\eta^2 = .028$  (see Table 1). The magnitude of rating differences between German and Turkish stimuli was significantly related to open prejudice expressions in the questionnaires, with  $r = .16$ ,  $p < .001$  for the blatant scale and  $r = .32$ ,  $p < .001$  for the xenophobia scale.

#### *Priming effects*

Analyses of priming effects were based on error rates throughout the experimental phase (200 trials). Mean error rate was 23.75 % ( $SD = 9.60$  %). The mean response latency was 484 ms ( $SD = 58$ ). In correspondence to the response deadline technique, no systematic influences of prime presentation on response latencies were found,  $F_s < 2$ ,  $ps > .15$ . Priming indices were calculated by subtracting the mean error rates in congruent trials from incongruent trials, thus representing the crucial interaction of prime valence and target evaluation (see Study 1). A significant priming effect was found in the standard priming condition ( $M = 2.82$  %,  $SD = 8.52$ ),  $t(263) = 5.37$ ,  $p < .001$ ,  $d = .33$ . More importantly, age did not affect the standard priming score,  $F(3,260) = 0.19$ ,  $ns$ . Significant priming effects of comparable size were found in all age groups (all  $t_s \geq 2.20$ ,  $ps \leq .03$ ;  $.30 \leq d \leq .34$ ). Thus, prime valence systematically influenced target evaluations, regardless of participants' age. It can therefore be concluded that the

masked priming task was equally sensitive for effects of automatic valence activation in children and adolescents, as compared to the priming procedure used in Study 1.

In the prejudice condition with Turkish and German primes, a significant effect of participants' age occurred,  $F(3,260) = 2.62, p = .05, \eta^2 = .029$  (linear trend:  $F(1,260) = 4.44, p = .03$ ). Only priming effects of eighth-graders differed significantly from zero,  $t(60) = 2.66, p = .02, d = .31$ ; no significant prejudice-related priming effects were found in any of the younger age groups, all  $ts \leq 1.17, ps \geq .25$  (see Figure 2). Planned contrasts (Helmert) comparing each age group with the respective younger age group(s) showed only one significant difference between priming effects of the eighth-graders and mean priming effects of the younger participants,  $F(1,260) = 7.72, p = .006$ , indicating a developmental shift between 7<sup>th</sup> and 8<sup>th</sup> grade. On the contrary, priming effects of 7<sup>th</sup> and 6<sup>th</sup> graders did not differ from the respective younger participants,  $F_s < 1, ns$ .

As in Study 1, separate priming indices were computed for German and Turkish primes, in order to explore whether the age effect was driven by reactions towards ingroup vs. outgroup members. As in Study 1, no significant effect of age was found for the German priming index,  $F < 1, ns, (M = -0.16, SD = 15.64)$ , whereas a significant effect of age was found for the Turkish priming index,  $F(3, 260) = 2.66, p = .05, \eta^2 = .03$  ( $M_{5th} = 1.94\%, SD = 15.65; M_{6th} = 2.80\%, SD = 14.52; M_{7th} = 0.91\%, SD = 15.42; M_{8th} = -4.10\%, SD = 12.57$ , respectively).

Similar to Study 1, internal consistency of the Turkish priming index (controlled for the German index, see Study 1 for procedural details) was low with a median

Spearman-Brown-adjusted split-half correlation of  $r = .31$ . As previously, priming effects were unrelated to questionnaire scores,  $-.02 \leq rs \leq .05$ , *ns*, and were also unrelated to explicit stimulus ratings,  $-.06 \leq rs \leq .06$ , *ns*. Again, these results indicate a dissociation of automatic valence activation and open attitude expression.

#### *Prime awareness*

None of the participants reported having noticed the hidden primes during the priming task. With the data of the direct test, we computed the non-parametric signal detection sensitivity index  $A'$  for the categorizations of the masked primes. Indices were computed separately for the positive-vs.-negative detection task and the Turk-vs.-German-detection task. Mean  $A'$  was  $M = .51$  ( $SD = .16$ ) for the positive-negative detection,  $t(263) = 1.09$ , *ns*, and  $M = .54$  ( $SD = .16$ ) for the Turkish-German detection,  $t(263) = 3.67$ ,  $p < .001$ , both tests concerning the deviation from chance level .50. Thus, whereas the detection of standard primes did not differ significantly from chance, the direct test yielded an above-chance result for the recognition of Turkish and German primes. Detection of standard primes was unrelated to age,  $F(3,260) = .28$ , *ns*, but performance in the Turkish-German classification test increased with age,  $F(3,260) = 2.83$ ,  $p = .04$ . Whereas fifth- to seventh-graders showed detection indices that did not differ significantly from chance level ( $M_{5th} = .50$ ,  $SD = .16$ ,  $M_{6th} = .53$ ,  $SD = .17$ ,  $M_{7th} = .53$ ,  $SD = .17$ ,  $ts < 1.5$ , *ns*), eighth -graders showed somewhat better discrimination ( $M_{8th} = .58$ ,  $SD = .16$ ,  $t(60) = 4.03$ ,  $p < .001$ ). Thus, although none of the participants reported conscious awareness of prime presentation during the affective priming task,

older participants were somewhat better able to discriminate German from Turkish primes when specifically instructed to do so.

We took several measures to ensure that the priming effects of eighth-graders cannot be attributed to a higher prime sensitivity, as indicated by the direct test. First of all, the prejudice-related priming effect was not related to prime sensitivity in the direct test ( $r = -.05$ , *ns*), thus prime sensitivity cannot be a mediator of the age effect on priming. Correspondingly, in an analysis of covariance using the Turk/German prime sensitivity as a covariate, the age effect on priming remained significant,  $F(3,260) = 3.03$ ,  $p = .03$ ,  $\eta^2 = .034$ , and no significant effect of prime sensitivity occurred,  $F < 1$ , *ns*. Second, to show that the priming effect found for eighth-graders did not depend on those children with partial awareness, we followed the suggestion by Draine and Greenwald (1998) to regress the (prejudice-related) affective priming index on a transformation of  $A'$  centered around zero (i.e., we subtracted 0.5 from the original  $A'$  for each participant). While the centered  $A'$  is associated with a non-significant regression weight,  $t(59) = -0.09$ , *ns*, the intercept of this regression deviated significantly from zero with a regression weight of  $B = 2.61$ ,  $t(59) = 2.13$ ,  $p = .04$ . Since both the priming index and the direct measure have a natural zero point, we can interpret this effect as an above-zero priming effect in the absence of objective prime awareness.

### *Discussion*

Results of this study show that the affective priming task is applicable for the assessment of spontaneous evaluations in children and adolescents, even when the

primes are presented outside conscious awareness. The significant standard priming effects with IAPS-pictures as primes demonstrate that this technique and the non-masked priming procedure used in Study 1 are similarly sensitive to automatic evaluation.

Most importantly, we replicated the age influence upon the prejudice-related priming effect that had been observed in Study 1. Whereas adolescents attending eighth grade showed significant, prejudice-related priming effects, no such effects occurred for younger children in the fifth to seventh grades. Note that these priming effects occurred independently from participants' prime awareness.

The age effects on open prejudice expression in the questionnaires of Study 1 could not be replicated in Study 2. Participants of all age groups expressed relatively low levels of open prejudice against Turks or general xenophobia. The ratings of liking for prime portraits showed a comparable age effect, however: Participants of all age groups expressed levels of liking for the German pictures that were significantly higher than for Turkish pictures. This difference in liking for Turkish and German pictures decreased with higher age. Open prejudice expression was again unrelated to affective priming effects, suggesting a dissociation of the developmental course of open prejudice expression and automatic activation of prejudice-related evaluations.

The age differences found in the affective priming effect for Studies 1 and 2 stand in contrast to results of previous IAT studies, in which no effects of age on automatic prejudice were found (e.g., Baron & Banaji, 2006; Dunham, Baron, & Banaji, 2006; Rutland et al., 2005). We assume that this contrast is related to the different

measurement methods – affective priming vs. IAT – and the way they assess prejudice-related evaluations. However, before elaborating on this assumption and its implications for developmental theories of prejudice automatization, at least one further study is needed that directly demonstrates that the age effects of Study 1 and 2 are related to the specific measurement procedure of affective priming. Thus, in Study 3, we aimed to rule out alternative explanations, based on differences between previous IAT studies and the current studies related to the societal and cultural context of the experiments. Previous IAT studies have been conducted with Blacks and Whites in the US (Baron & Banaji, 2006) or the UK (Rutland et al., 2005) as target groups. These groups differ from Turks in Germany in many regards, for example in the degree of ‘visibility’ or perceptual salience compared to the majority ingroup, and in historical as well as contemporary characteristics of the specific intergroup relations. If the age effects found in Study 1 and 2 were related to such differences, the same age effect should be found using any measure of automatic prejudice activation, whether it is an affective priming task or an IAT. If the age effect is related to the measurement procedure of affective priming, as we assume, it should not be found with an IAT.

### Study 3

This study aimed to replicate the age effects on automatic prejudice activation, found with the affective priming tasks in Studies 1 and 2. Furthermore, we explored whether a similar age effect would be found in a prejudice IAT, using pictures of Turks and Germans as target categories. Participants in Study 3 therefore completed both an affective priming task and an IAT, using the same set of stimuli for the assessment of

automatic prejudice. Furthermore, we expanded the age range to also include ninth-graders.

### *Method*

#### *Participants*

Participants were 141 students from grades five to nine, drawn from a different public high school (situated in the same region). The sample was comprised of 43 fifth-graders (22 boys, 21 girls; median age = 10, ranging from 9 to 11), 31 sixth-graders (12 boys, 19 girls; median age = 11, ranging from 11 to 12), 26 seventh-graders (9 boys, 17 girls; median age = 12, ranging from 11 to 14), 16 eighth-graders (7 boys, 9 girls; median age = 13, ranging from 13 to 15) and 25 ninth-graders (13 boys, 12 girls; median age = 14, ranging from 13 to 15). The data of 29 additional participants with immigration backgrounds were discarded from analyses. Data of five more participants were incomplete, due to computer failure. Furthermore, for 14 fifth-grade participants, no IAT were recorded due to computer malfunction. Similar to Study 2, data were collected at the beginning of the school year.

#### *Materials and Procedure*

The affective priming task contained the same priming stimuli as in Study 1 and 2. For the IAT, the same 10 German and 10 Turkish prime pictures represented the target categories, and a selection of 10 positive and 10 negative target pictures from the priming task represented the so-called attribute dimension in the IAT.

Children always first completed the affective priming task, followed by the IAT. The affective priming procedure was identical to that of Study 1; with the exceptions

that (a) the SOA was shortened to 317 ms, by immediately replacing the prime stimulus by the target stimulus, and (b) no additional memory task instructions were given. The primes were introduced as “distracter pictures”, and children were explicitly instructed to ignore them and focus on the target evaluation task.

After the priming task, participants took a short break and subsequently worked through the IAT. They were first familiarized with the classification task for Turkish and German pictures. They were told that the pictures which had served as distracters in the precedent task were now task-relevant, and would have to be categorized as depicting either Turkish or German young men. The Turkish-German categorization task was practiced in 20 trials. Response key assignment was counter-balanced between groups of participants. As instructions were given verbally, children in one experimental session were always included in the same condition. No additional practice phase was administered for the positive-negative categorization task, as it was identical to the evaluation task children had completed in the priming task. Instead, children were informed that the two tasks were now mixed, and that they had to categorize the colored pictures as being positive or negative, and the black and white portrait pictures as depicting either a Turkish or a German person. Participants then started immediately with the first mixed block of the IAT, in which target and attribute stimuli were presented in a random, alternating order. According to category-key assignment, approximately half of the participants started the IAT with a congruent vs. incongruent condition, respectively. The first block consisted of 50 trials, of which the first 10 trials were regarded as additional practice trials. Afterwards, participants were familiarized

with the reversed key-assignment for the Turkish and German trials and practiced it in another 20 practice trials. They then continued with the second mixed block, which again consisted of 10 additional practice trials and 40 experimental trials. Participants took another short break after completion of the IAT and before the liking ratings were obtained for all Turkish or German portraits. Because of time constraints, no further explicit prejudice measures were obtained.

### *Results*

#### *Stimulus ratings*

The pleasantness ratings of the portrait pictures of Turks and Germans were submitted to a 2 (prime origin: Turkish vs. German) by 5 (participant's age: 5<sup>th</sup> vs. 6<sup>th</sup> vs. 7<sup>th</sup> vs. 8<sup>th</sup> vs. 9<sup>th</sup> grade) analysis of variance. A significant main effect of prime origin on the ratings was found,  $F(1,136) = 98.75, p < .001, \eta^2 = .421$ , indicating that German stimuli ( $M = 4.57, SD = .87$ ) were rated significantly higher in pleasantness than Turkish stimuli ( $M = 3.38, SD = 1.31$ ). No other effect gained significance, indicating that this bias was unaffected by participants' age,  $F < 1, ns$  (see Table 1).

#### *Priming effects*

Priming effects were calculated as described above, based on trimmed response latencies of correct response throughout the experimental phase (see Study 1). Mean error rate was 7.41 % ( $SD = 5.60$  %).<sup>5</sup>

A significant priming effect was found in the *standard priming* condition,  $M = 18$  ms ( $SD = 29, t(141) = 8.25, p < .001, d = 0.70$ ). As in the previous studies, age did not

affect the priming score,  $F(4,136) = 1.56, p = .19$ . Significant priming effects of comparable size were found in all age groups (all  $t_s \geq 2.30, p_s \leq .03; 0.42 \leq d_s \leq 0.96$ ).

A significant effect of participants' age was found in the *prejudice priming* condition with Turkish and German primes,  $F(4,136) = 2.50, p = .04, \eta^2 = .07$  ( $F[1,136] = 6.77, p = .01$  for the linear trend). In accordance with the previous studies, only priming effects of eighth-graders ( $M = 12$  ms,  $SD = 16, t(15) = 3.07, p = .008, d = 0.74$ ) and ninth-graders ( $M = 11$  ms,  $SD = 16, t(24) = 3.41, p = .002, d = 0.68$ ) differed significantly from zero, whereas no significant prejudice-related priming effects were found for any of the younger age groups, all  $t_s \leq 1.66, p_s \geq .11, -0.01 \leq d_s \leq 0.32$  (see Figure 3). Planned contrasts between age groups (Helmert) were computed as in the previous studies. In Study 3, however, the contrast indicating a developmental shift between 8<sup>th</sup> grade and the younger grades, missed significance,  $F(1,136) = 2.49, p = .12$ . However, the difference between sixth- and fifth-graders was marginally significant,  $F(1,136) = 2.84, p = .09$  (all other  $F$ 's  $< 1.63, ns$ ).

Separate priming indices were computed for German and Turkish primes. Again, no significant effect of age was found for the German priming index,  $F < 1, ns, (M = 31$  ms,  $SD = 38)$ , whereas a significant effect of age was found for the Turkish priming index,  $F(4,136) = 3.68, p = .007, \eta^2 = .098$  ( $M_{5th} = 39$  ms,  $SD = 51; M_{6th} = 21$  ms,  $SD = 36; M_{7th} = 18$  ms,  $SD = 31; M_{8th} = 16$  ms,  $SD = 28; M_{9th} = 3$  ms,  $SD = 30$ , respectively). As in the previous studies, internal consistency of priming effects was rather low with a median Spearman-Brown-adjusted split-half correlation of  $r = .38$  (see Study 1 for

procedural details). Separate priming effects, as well as composite scores, were unrelated to stimulus ratings,  $-.04 \leq r_s \leq .07$ , *ns*.

#### *IAT scores*

For reasons of comparability, RT data in the IAT were trimmed according the same criteria used for the affective priming task, excluding 0.8% of the trials. Analyses were conducted only for trials with correct responses. The mean error rate was 7.74% ( $SD = 5.93$ ). IAT scores were computed as the difference between compatible and incompatible blocks (with compatibility defined from the perspective of positive ingroup and negative outgroup evaluation). Additional analyses were conducted based on the IAT scoring algorithm (D) recommended by Greenwald, Nosek, and Banaji (2003) that yielded the same pattern of results. Overall, a significant IAT effect was found, ( $M = 89$  ms,  $SD = 108$ ;  $t(131) = 10.15$ ,  $p < .001$ ,  $d = 0.88$ ;  $D = 0.44$ ,  $SD = 0.37$ ,  $t(131) = 13.45$ ,  $p < .001$ ). The IAT score was submitted to a 2 (block sequence: compatible first vs. incompatible first) by 5 (participant's age: 5<sup>th</sup> vs. 6<sup>th</sup> vs. 7<sup>th</sup> vs. 8<sup>th</sup> vs. 9<sup>th</sup> grade) analysis of variance. This analysis revealed a significant main effect of block order,  $F(1, 131) = 6.15$ ,  $p = .01$ ,  $\eta^2 = .05$ , indicating that IAT effects were significantly enlarged when the task started with the compatible block, ( $M = 107$  ms,  $SD = 119$ ;  $D = .53$ ,  $SD = .37$ ), as compared to the incompatible block ( $M = 65$  ms,  $SD = 85$ ;  $D = .32$ ,  $SD = .35$ ). Most importantly, no significant main effect of age was found, nor was a significant interaction effect of age with block sequence,  $F_s < 1$ , *ns*, replicating results of other studies. Significant IAT scores occurred in all age groups with  $M_{5th} = 90$  ms ( $SD = 138$ ),  $M_{6th} = 104$  ms ( $SD = 75$ ),  $M_{7th} = 66$  ms ( $SD = 86$ ),  $M_{8th} = 104$  ms ( $SD =$

144),  $M_{9th} = 84$  ms ( $SD = 91$ ),  $t_s > 2.88$ ,  $p_s \leq .01$ . The Spearman-Brown-adjusted reliability score for the IAT (D score) was  $r = .73$ . The IAT score was unrelated to the prejudice-related affective priming score,  $r = -.09$ , *ns*. However, the IAT score was negatively related to explicit evaluations of the Turkish stimuli, with  $r = -.25$ ,  $p < .01$  but no significant relation was found between the IAT score and the mean ratings of German stimuli,  $r = .004$ , *ns*. Thus, the higher the IAT score, the more negative the evaluations of the Turkish stimuli.

For a direct comparison of age differences on prejudice effects found with the different measures, z-transformed IAT and priming scores were submitted to a 2 (Measure: IAT vs. Priming) by 5 (Participant's age) mixed model ANOVA which revealed a significant interaction when testing for linear trends,  $F(1,128) = 4.74$ ,  $p = .03$ . Thus, differences between the age effects on the different measures can be interpreted as such.

### *Discussion*

In Study 3, the age effect on prejudice-related priming found in the previous studies was replicated. Again, significant priming effects indicating relative negativity towards Turks as compared to Germans were only found in children older than 13 years of age. Most importantly, no such age effect was found in an IAT using the same stimuli. Rather, a significant prejudice effect in this task was found throughout the age range replicating IAT results found in other prejudice (e.g., Baron & Banaji, 2006; Rutland et al., 2005). This pattern of results indicates that the age differences found in our studies are unlikely to be related to cultural differences or peculiarities of the current sample.

Rather, the contrast between results of our studies and previous studies on automatic prejudice should be explained in terms of the characteristics of the two measures, affective priming and IAT.

One of the most blatant differences between the two measures is the salience and task-relevance of social categorizations: In the IAT, prejudice assessment is based on forced-choice categorizations of target stimuli, and IAT scores thus represent evaluations of a given *category label*. One cannot straightforwardly conclude that the categorizations predetermined in the IAT are actually the ones children spontaneously use in person perception. Accordingly, one cannot conclude that that individual category members are spontaneously evaluated the same way as the category labels (see De Houwer, 2001; Olson & Fazio, 2003 for related arguments). In the affective priming paradigm, on the contrary, category-based evaluations play a subordinate role, and the measure is expected to assess spontaneous evaluations on an *exemplar-based level* (Olson & Fazio, 2003; Livingston & Brewer, 2002). We assume that the contrast between stable category-based prejudice effects in the IAT and reliable age differences on exemplar-based prejudice in the affective priming task reveals differences in the developmental processes underlying prejudice activation, related to the degree of spontaneous vs. forced category use.

It is well demonstrated that children acquire awareness of racial and ethnic categories and their societal evaluations very early, between the ages of three to six years (e.g., Aboud, 1988, 2008). These evaluations of social categories might become stably represented in long-term memory and will be recalled when children are

explicitly prompted to categorize people according their social group memberships, such as in the IAT (and any other forced choice measure). This, however, does not necessarily imply that in real-life encounters children actually (a) do *automatically* categorize people according their ethnicity, and that (b) they do *automatically* activate the corresponding, category-based evaluations. Children might be able to indicate whether a person is black or white, or German or Turkish, and might also express societal intergroup biases if explicitly asked to do so. However, such categorizations and evaluations might not yet play a role in children's spontaneous person perceptions. For example, it has been repeatedly demonstrated that young children are familiar with societal evaluations of specific social category labels without being able to correctly assigning the labels to specific category members (e.g., Bar-Tal & Teichman, 2005; Nesdale, 2001). Thus, the finding of stable IAT effects in response to social category labels (Baron & Banaji, 2006; Rutland et al., 2005) does not imply that children automatically activate negative evaluations in encounters with outgroup members whose category membership is not made salient. Further, the notion of unconditional automatic categorization has also been questioned for adults (e.g., Macrae, Quinn, Mason, & Quadflieg, 2005). It thus makes sense to assume a developmental change from early category-based prejudice expression to later exemplar-based prejudice activation: Whereas children are early familiarized with societal evaluations of social categories that they reproduce if category membership is made salient / task relevant (as documented by the IAT), they only later acquire the necessary expertise for effortlessly applying social categories to individuals and thus show spontaneous activations of

category-related evaluations only later in life (as documented in the affective priming paradigm). These assumptions overlap with theories on ethnic identity development (e.g. e.g., Quintana, 1998) or theories of prejudice development (e.g., Nesdale, 1999, 2001) that describe age-related increases in the use of categorical information in person perception and evaluation. Other developmental models, like the social cognitive developmental theory by Aboud (1988, 2008), however assume a decrease in category reliance in children above the age of seven – related to their increasing cognitive abilities. The goal of Study 4 was thus to test (a) whether the assumed age-related increase in the spontaneous use of ethnicity can be found and (b) which role spontaneous vs. forced categorizations play for the automatic activation of intergroup prejudice in the same age range.

#### Study 4

To test these assumptions, a further study was conducted following two complementary strategies. First, we aimed at demonstrating that the affective priming paradigm yields the same results with children and adolescents as the IAT if social categorizations are made similarly task relevant as in the IAT. In Study 1, we had made an unsuccessful attempt to test for category based vs. exemplar based prejudice activation within the affective priming task, following Olson and Fazio (2003). However, this manipulation – linked to an additional prime memorization task – seems to have been too subtle to actually alter children's prime processing as they tended to ignore or forget this additional task. We therefore used a stronger manipulation in Study 4 in which social categorizations were made task relevant. Children had to indicate the

social category membership of prime persons (Dutch vs. Moroccan) immediately after each target reaction (*Forced Categorization* condition). With this task, prime salience was kept constantly high and prime categorization was task relevant as in a typical IAT. The remaining participants completed a standard affective priming procedure as described in Study 3 (*No Categorization* condition). We expected affective priming effects in the forced categorization condition to be similar to IAT effects of Study 3, thus showing a stable prejudice-related priming effect independent of participants' age. In the no categorization condition, the typical age differences found in Study 1 to 3 should be replicated.

Second, we aimed at testing directly whether age-related differences in children's spontaneous use of social categories in person perception could be found, implementing a measure of spontaneous category use, the Who-said-What paradigm (Taylor, Fiske, Etoff, & Ruderman, 1978, see Bennett & Sani, 2003, for use of the paradigm with children). In this paradigm, children are shown a series of individuals making a number of statements (e.g., "In the afternoon, I like to play soccer with my friends."). Each individual is identified by a photograph. In our adaption, the target individuals orthogonally varied in terms of age (child vs. adult) and ethnicity (white Dutch vs. Moroccan). Most importantly, neither age nor ethnicity was explicitly mentioned throughout the task and the content of the statements was unrelated to social category memberships. Children were later asked in a surprise recall test to remember the statements and the persons who have made them. The dependent variable of interest in this paradigm is the number and distribution of identification errors, thus cases in which

statements are assigned to the wrong person: If within-category confusions occur more frequently than between category confusions, it can be inferred that the category membership has been attended to, was encoded in memory, and subsequently used in the memory task. Thus, this paradigm allows indirect assessment of spontaneous use of social categories and testing whether age-related increases in spontaneous category use mediates age effects on prejudice related priming scores. Furthermore, by orthogonally manipulating two social categories (i.e., age and ethnicity) the paradigm allows estimation of age effects in the relative salience of these two categories.<sup>6</sup>

For the current research, we adjusted the classical procedure according to suggestions by Klauer and Wegener (1998). In the classical paradigm, participants do not have the option to state that they do not remember the statement and have to guess the speaker. This forced guessing invites for possible confounds – especially because one can assume that memory performance differs with participants' age and younger participants are less likely to remember a statement as compared to older children. For example, if guessing is independent of category membership, poor statement memory and frequent guessing would level out differences between within-category and between-category errors and thus the spontaneous use of category information would be underestimated for (younger) children with poor statement performance. For this reason, we adjusted the memory test by interspersing new statements. Children indicated first whether they remembered a statement as having been presented before (old) or not (new). Only for as old recognized statements they were prompted to indicate who had given it (see Klauer & Wegener, 1998).

Study 4 was conducted in the Netherlands; contrasting attitudes towards the white Dutch majority ingroup and the Moroccan immigrant outgroup. Thus, the study took place in a different socio-cultural and intergroup context and therefore also explored the generalizability of results of the first three studies to an intergroup context outside Germany.

### *Method*

#### *Participants*

Participants were 129 students from grades five to eight of a public Dutch elementary school, situated in a small-sized town in a Central Netherlands region with slightly lower percentages of (non-western) immigrants as in the previous studies (approx. 5 %). Approximately 2% of the school populations were from a visible ethnic minority, all teachers were white Dutch. The sample is divided into 29 fifth-graders (15 boys, 13 girls, 1 not reported; median age = 9), 30 sixth-graders (18 boys, 12 girls; median age = 10), 36 seventh-graders (19 boys, 17 girls; median age = 11), and 34 eighth-graders (15 boys, 18 girls, 1 not reported; median age = 12). Prior to the study, we obtained written parental consent for the child's participation. Incomplete data of 26 further participants (due to computer malfunctions or individual circumstances impeding data collection) as well as data of 5 further students with immigration backgrounds were discarded from analyses. The study was conducted in the computer classroom of the schools at the end of the school year prior to the summer vacation. In the Netherlands, mandatory education starts at the age of five, which explains why

students of the same grades were one year younger, as compared to the German students in the previous studies who start school at the age of 6 years.

### *Materials*

*Priming measure.* The priming task conformed to a 2 x 4 within-subjects design, resulting from the factorial combination of target valence (positive vs. negative) and prime type (positive vs. negative vs. Moroccan vs. Dutch). A new set of portraits of 20 young men of white Dutch vs. Moroccan origin was selected from a database of 280 portrait pictures after extensive pretesting with Dutch students. The selected Dutch and Moroccan pictures were matched according to perceived ingroup typicality, age-estimates, attractiveness and emotionality of facial expression. Contrary to the previous studies, colored pictures were used which slightly increases perceptual salience of group membership. All faces were in frontal view against a grey background, and approx. 75 mm high and 50 mm wide. The IAPS pictures used in previous studies were included as standard primes and targets. Contrary to the previous studies, pictures depicting persons were not used as targets (to avoid that children mistakenly apply the forced categorization task to the target instead of the primes).

The “Who-said-What” paradigm conformed to a 2 x 2 within-subjects design, resulting from the factorial combination of target age (child vs. adult) and target ethnicity (Moroccan vs. Dutch). Therefore, portraits of four adults and four children of white Dutch vs. Moroccan origin were selected from the same database according the same matching criteria (see above). For the memory task, two sets of 24 statements were created that describe a person’s preferences for food, activities or pets (e.g., “*In the*

*afternoon, I like to play soccer with my friends.”*). One set was assigned to be ‘old’, to be presented during the learning face, the other to be ‘new’, to be presented as filler items in the memory test.

### *Procedure*

Participants were tested in groups of 6 to 8 students seated separately in computer classrooms of their school. All instructions were given by one of two white Dutch female experimenters. The experimental session always started with the Who-said-What paradigm. Children were informed that they would get to know a group of eight people who tell them about their favorites and that their task was to make an impression of these people. Presentation followed a fixed random order in which each statement was presented together with a picture for 8 seconds. Children were instructed to read the statement and look at the person. In the fifth and second grade, the children were instructed to read the statements aloud together with the experimenter. After viewing the 8 individuals each with 3 statements, children were introduced to the surprise recall task in which they had to remember the statement and which target person had made it. Therefore, the original 24 statements were interspersed with the 24 new statements. Each statement was presented one by one on the computer screen and children first indicated if they had previously encountered the statement or not via mouse click on an ‘old’ or a ‘new’ button. If the statement was classified as old, the pictures of all eight target persons were subsequently presented on screen (in one fixed random order) and participants chose the person that they thought had made the statement previously. If

they classified a statement as new, the task continued directly with the next statement.

Completion of the Who-said-What paradigm lasted about 5 minutes.

Before the affective priming procedure started, all children completed two practice blocks in which each prime picture had to be categorized according their social group membership<sup>7</sup>. To keep the procedure identical with the later priming trials, each picture was preceded by a fixation cross presented for 1000 ms. The picture itself was presented for 316 ms and than replaced by a blank. Finally, with a delay of 1200 ms, a probe question appeared, asking whether the previous picture depicted a Dutch person or whether it depicted a Moroccan person (i.e., “Was dat een NEDERLANDER?” or “Was dat een MAROKKAN?”). Children responded by pressing a right key [M] for ‘yes’ and a left key [C] for ‘No’. Key-response reminders were presented in the lower left and right corners of the screen. They received instantaneous performance feedback, with the messages “JUIST!” (correct) or “FOUT!” (incorrect) appearing on screen for 400 ms. Throughout the experiment the type of question was counterbalanced, such that Children had to respond equally often with ‘yes’ and ‘no’ to Dutch and Moroccan primes. In the second practice block, the standard primes were interspersed into the prejudice-related primes. In these trials, no response probe appeared and children were instructed to ignore these pictures and wait for the next portrait picture. After completion of the categorization practice blocks, children were familiarized with the target evaluation task, in which they had to respond as quick as possible to positive and negative target pictures by pressing the right key [M] for positive and the left key [C]

for negative pictures. Key-response reminders were presented in the lower left and right corners of the screen. The task was practiced in 16 trials.

After this practice phase, the affective priming task started with the between participants manipulation of prime categorization. In the *no categorization* condition, priming trials followed the procedure of Study 3. That is, a trial started with the fixation cross presented for 1000 ms, followed by the prime presented for 316 ms, directly followed by the target. If participants did not respond within 1750 ms, a message appeared on screen reminding them to react faster. Children were instructed to ignore the presentation of the first picture (the prime) and focus on evaluating the target as fast as possible. In the *forced categorization* condition, children were instructed that they had to complete two intermixed tasks, that is they had to react first to the target picture by evaluating it as fast as possible and were then asked to categorize the preceding prime picture. The trial timing was identical with the no-categorization condition with the following exception: After children pressed the key in the target evaluation task, one of the two prime categorization probes appeared (randomly selected) asking whether the preceding portrait depicted a Dutch person or whether it depicted a Moroccan person. The next trial started with a delay of 500 ms after a response was recorded. The priming task was practiced in 24 further trials. The main phase consisted of 192 trials separated into 3 blocks. Throughout the experiment, children received instantaneous accuracy feedback. Furthermore, overall performance feedback on average speed and accuracy in the evaluation task was presented at the end of each block. Children in the *forced categorization* condition yielded additional block-wise feedback on accuracy in the

prime categorization task. After completion of the priming task, children filled out a computerized demographic questionnaire and were then debriefed by the experimenter. Completion of the study lasted approximately 30 minutes in the no-categorization condition and 40 minutes in the forced-categorization condition.

### *Results*

*Priming effects.* Priming effects were calculated as described above, based on trimmed response latencies of correct response throughout the experimental phase (see Study 1). Mean error rate in the evaluation task were 5.22 % ( $SD = 7.20$  %) and 6.30 % ( $SD = 7.57$  %) for the no-categorization and the forced-categorization conditions, respectively.

In the in the *standard priming* condition, a significant priming effect was found with  $M = 26$  ms ( $SD = 45$ ,  $t(128) = 7.91$ ,  $p < .001$ ,  $d = 0.70$ ). As in the previous studies, age did not affect the priming score,  $F(3,121) = 1.72$ ,  $p = .17$ . Also, the categorization condition had no significant effects on the standard priming score,  $F < 1$ , *ns*.

Analysis of *prejudice priming* effects revealed a significant main effect of condition,  $F(1,121) = 4.27$ ,  $p = .04$ ,  $\eta^2 = .034$ . The main effect of grade did not reach significance,  $F(3,121) = 1.49$ ,  $p = .23$ ,  $\eta^2 = .035$ , nor did the interaction of grade and condition,  $F(3,121) = 1.09$ ,  $p = .36$ ,  $\eta^2 = .026$ . However, testing for the linear trend, we found a significant interaction of grade and condition,  $F(1,121) = 2.81$ ,  $p = .05$  (one-tailed). As can be seen in Figure 4, no significant age effect was found in the *forced categorization* condition,  $F < 1$ , *ns* with an overall significant affective priming effect of  $M = 20$  ms ( $SD = 62$ ),  $t(61) = 2.46$ ,  $p = .017$ ,  $d = .34$ . In contrast, priming effects in the

*no categorization* condition were significantly affected by participants age,  $F(3,64) = 3.05, p = .035, \eta^2 = .125$  (with  $F(1, 64) = 8.43, p = .005$  for the linear trend). In accordance with the previous studies, only priming effects of eighth-graders ( $M = 15$  ms,  $SD = 21, t(14) = 2.81, p = .014, d = 0.72$ ) differed significantly from zero, whereas no significant prejudice-related priming effects were found for any of the younger age groups, with  $ts \leq .20, ps \geq .23$ , for the sixth and seventh grade and a even marginally significant reversed score for the fifth graders  $t(15) = 1.95, p = .70, d = -.48$  (see Figure 4). Planned contrasts between age groups (Helmert) indicated a significant developmental shift between 7<sup>th</sup> and 8<sup>th</sup> grade (i.e., the mean of the 8<sup>th</sup> grade differed from the average mean of the subsequent grades),  $F(1,64) = 3.14, p = .04$  (one-tailed), whereas contrasts for the 7<sup>th</sup> grade (vs. 6<sup>th</sup>/5<sup>th</sup> grade) and 6<sup>th</sup> grade (vs. 5<sup>th</sup> grade) were not significant,  $F_s < 1.19, ps > .28$ .

Separate priming indices were computed for Dutch and Moroccan primes to explore the role of ingroup positivity vs. outgroup negativity for the age effect in the *no categorization* condition. The pattern of results slightly differed from the previous studies in that the Dutch (ingroup) priming effect and the Moroccan (outgroup) priming effect were both affected by age in the opposite direction: Whereas a linearly increasing trend was found for the Dutch priming scores,  $F(3, 64) = 2.14, p = .15$  (with  $M_{5th} = -3$  ms,  $SD = 63; M_{6th} = 19$  ms,  $SD = 67; M_{7th} = 35$  ms,  $SD = 76; M_{8th} = 27$  ms,  $SD = 37$ , respectively), a linearly decrease was found for the Moroccan priming effect,  $F(3, 64) = 3.93, p = .05$  (with  $M_{5th} = 51$  ms,  $SD = 73; M_{6th} = 23$  ms,  $SD = 53; M_{7th} = 40$  ms,  $SD = 72; M_{8th} = -2$  ms,  $SD = 46$ , respectively). This pattern of results indicates that

developmental changes in prejudice activation were related to both, increasing ingroup positivity and increasing outgroup negativity. As in the previous studies, priming effects were characterized by rather low internal consistencies, with median Spearman-Brown-adjusted split-half correlation of  $r = .32$  for the forced-categorization condition and  $r = .36$  for the no-categorization condition (see Study 1 for procedural details).

*Who-Said-What Paradigm.* Of the forty-eight statements, children recognized 18.7 ( $SD = 2.9$ ) statements correctly as old and classified 22.5 ( $SD = 2.5$ ) statements correctly as new. Furthermore, 51.96% ( $SD = 18.12$ ) of the correctly as old classified statements were assigned to the correct speaker. Whereas no significant effect of age was found in the old-new discrimination,  $F < 1$ , *ns*, a significant age effect was found for the speaker assignment,  $F(3, 119) = 7.69$ ,  $p < .001$ ,  $\eta^2 = .162$ . As expected, older children showed a higher proportion of correct assignments compared to younger children (see Table 2). For analysis of errors, we computed the rate of within-age-within-ethnicity, within-age-between-ethnicity, between-age-within-ethnicity and between-age-between-ethnicity errors for statements that were correctly perceived as old. As the number of between-category-confusions expected by chance is two times greater than the number of within category confusions (for any given target there is only one other target within the combined age-ethnicity subcategory but two targets in each of the other categories), we corrected between-category errors by multiplying with 2 (see Stangor, Lynch, Duan, & Glass, 1992). The resulting error rates were subjected to a 2 (age error type: within vs. between) by 2 (ethnicity error: within vs. between) by 4 (grade) analysis of variance with repeated measurement on the first two factors. This

analyses revealed a significant main effect of age error type,  $F(1,119) = 281.85, p < .001, \eta^2 = .703$ , as well as a significant main effect of ethnicity error type,  $F(1,119) = 79.29, p < .001, \eta^2 = .400$ , that were qualified by their mutual interaction,  $F(1,119) = 73.96, p < .001, \eta^2 = .383$ . These results replicate findings from adult studies indicating a high proportion of within-age-within-ethnicity errors that indicate subtyping (see Stangor et al., 1992). The (adjusted) rate of these errors ( $M = 34.04\%$ ) was in all age groups higher than any other error type,  $ts > 3.42, ps < .003$ . Furthermore, a significant main effect of participants' age was found,  $F(3,119) = 4.76, p = .004, \eta^2 = .107$ , reflecting the aforementioned age effect in overall accuracy. Interestingly, a significant interaction of ethnicity error and participants' age was found,  $F(3,119) = 3.54, p = .017, \eta^2 = .082$ , that did not occur for the age-bias,  $F < 1, ns$ . For further analyses, we computed a composite score for ethnicity by subtracting mean between-ethnicity error rates from mean within-ethnicity error rates. This score significantly increased with children's age, from 10.22% (17.54) and 14.34% (26.66) in fifth and sixth grade to 27.12% (26.51) and 24.68% (21.99) in seventh and eighth grades, respectively. Use of ethnicity in the Who-said-What paradigm was significantly related to affective priming scores in the no-categorization condition,  $r = .24, p = .05$ , whereas this correlation missed significance in the forced-categorization condition,  $r = .22, p = .11$ .<sup>8</sup>

### Discussion

Study 4 was conducted with two goals. First, we tested whether prejudice related affective priming scores would show the same effect pattern as the IAT when ethnic categorizations were made salient and task relevant. Results show that this is indeed the

case: In the *forced categorization* condition of Study 4, significant priming effects were found that did not differ between age groups, replicating previous results of IAT studies. On the contrary, affective priming scores in the *no categorization* condition replicated results of the further studies, in that a significant linear trend of age was found and only priming scores of the eighth-graders were significantly different from zero. This finding also demonstrates that the age effects found in the previous studies were not bound to the social-cultural peculiarities of Germans' prejudice towards Turkish immigrants but can be generalized to a broader context of intergroup relations.

Furthermore, this study allowed direct testing of the suggestion that the age differences in prejudice-related priming scores (in the no-categorization condition) are related to children's increased spontaneous use of social categories in person perception. We therefore implemented a child-friendly version of the Who-said-What paradigm which allows indirect testing of spontaneous categorization use. Results indicate that children of all age groups use target age (child vs. adult) as well as target ethnicity (Dutch vs. Moroccan) when processing person information. Interestingly, whereas age was used to the same degree by participants of all age groups, an age-related increase in the use of ethnicity was found. This finding is consistent with our assumption that age-related differences in spontaneous categorizations influence automatic evaluation.

Thus, the increase of automatic prejudice activation with higher age can be (at least partly) explained by an increase of spontaneous category use in person perception. These results thus provide first support for the assumption that automatic prejudice activation develops in accordance with children's and adolescents' tendency to use

ethnicity as a socially significant category in person perception. Because the results of the mediation analysis (see Footnote 7) have to be interpreted with caution, further research is needed to further support this claim.

### General Discussion

To our knowledge, the current series of studies constitutes the first attempt to use an affective priming paradigm for the assessment of automatic prejudice activation in children and adolescents. The presented studies provided two important results: First of all, we have demonstrated that the affective priming task is generally applicable to the assessment of spontaneous evaluations in children and adolescents. When affect-loaded stimuli (i.e., positive and negative IAPS pictures) were presented as primes in the control conditions of all four studies, a typical affective priming effect occurred. Specifically, participants reacted faster and more accurately to targets of the same valence, and they reacted slower and less accurately to targets of a different valence. This result is very important in its own right, as it shows that children automatically evaluated affective stimuli in a similar manner to adults, even when primes were presented outside of conscious awareness as in Study 2. Furthermore, it demonstrates that the affective priming procedure can be used as a tool to explore developmental aspects of automatic evaluations and attitudes.

The second important result of this research is the finding of a stable age effect on automatic prejudice-related evaluations. In four studies, priming effects indicating relative negativity toward Turkish and Moroccan immigrants as compared to the German and Dutch ingroup were found only with young adolescents from the age of

12/13 onward. Children of younger age did not show any signs of different automatic evaluations toward ingroup vs. outgroup members. Separate analyses of priming scores illustrated that these age differences were mainly related to differential evaluations of outgroup pictures, whereas (positive) evaluations of ingroup pictures remained rather stable across the age groups. These results stand in sharp contrast to other studies using a different indirect measure of prejudice, the IAT (e.g., Baron & Banaji, 2006; Dunham et al., 2006; Rutland et al., 2005). In these studies, no age differences in prejudice-related prejudice scores were found, with effects being as strong at the age of 6 years as at the age of 18 years. Similarly, in our Study 3, no effect of age on prejudice scores could be found when the IAT was adapted to assess prejudice against Turks as compared to Germans.

The two measures, affective priming and IAT, differ on various characteristics; the most important in this context seems to be the task-relevance of social categorizations. Results of Study 4 demonstrated that a modified affective priming procedure, in which forced-choice social categorizations were made task relevant as in an IAT, yielded indeed prejudice effects similar to typical IAT studies with no age differences. This indicates that it is the forced use of social categorizations that prompts a direct activation of evaluations associated with these categories in children of all ages in the IAT as well as the modified affective priming procedure. However, if social categories are not made task relevant, such evaluations do not seem to be activated in young children below the age of 12/13 years. Results of Study 4 also show that the age-

related increase in affective priming effects in older children comes with an age-related increase of children's spontaneous category use.

The results of the current studies strongly suggest two different yet complementary developmental components of automatic prejudice in childhood and adolescence: An early onset of *category-based* prejudice automatization and a later onset of *exemplar-based* prejudice automatization. According to theories of ethnic identity development (e.g., Quintana, 1998; Hirschberg, 2008), children acquire the basic knowledge about the structure of their social environment at a very early age; they learn which groups exist, how they are labeled, and which evaluations are associated with them. Obviously, this knowledge is already well and stably represented in children's cognitive-affective systems, leading to fast and effortless activations of evaluative associations to social group labels as measured by an IAT (see Banaji et al., 2008). However, it seems that this early knowledge is quite abstract in the beginning. For example, young children lack an understanding of ethnicity below perceptual differences such as skin color (e.g., Quintana, 1998) and do not consistently rely on this categorical knowledge in their behavioral decisions (e.g., in playmate selection). Only in further development, children understand the social consequences of own and other's group memberships and learn to validly apply social categorizations to individuals. We assume that this understanding and the increasing use of social categorizations in daily life results in automatized activation of categorical evaluations in response to individual exemplars of social groups as measured by the affective priming paradigm.

The priming effects in the current studies followed linear age trends supporting this assumption. Additionally, we also found evidence for developmental shifts between the age of 12 and 13 years and affective priming effects indicating negative outgroup evaluations were only significant for the older participants. It remains to explain why such shifts would take place around this age and not at any other age. Is this age group characterized by an increased consideration of social categories that implies the above-stated frequent use? Indeed, research on social identity development points in that direction: Several authors have stated that identity development in early adolescence includes, as a critical experience, the formation of social identity (or identities). They have defined early- and middle-adolescence as a phase of increasing exploration of one's own social (ethnic) identity (e.g., French, Seidman, Allen, & Aber, 2006; Nesdale, 2001; Newman & Newman, 2002; Phinney & Chavira, 1992; Quintana, 1998). Of course, children are aware of being members of specific social groups or categories long before adolescence. However, group identity transforms into a dominant theme starting in late childhood and early adolescence. To understand own racial, ethnic, cultural, national, and sexual group memberships, and deliberately identify with own ingroups, differences to outgroups have to be explored, acknowledged and accentuated. Thus, ingroup vs. outgroup differentiations are sharpened, and social categories are used more frequently in person perception. Eventually, stable associations emerge between specific characteristics indicating group membership and related associations, which could explain why prejudice becomes automatically activated at this age. Of course, one cannot expect a generalized appearance of automatic prejudice at that age for all

children under all circumstances. On the contrary, Nesdale's (1999, 2001) social identity development theory assumes that the individual perception of ingroup attitudes, norms and values towards outgroups, as well as perceptions of intergroup relations, especially intergroup threat and conflict, play a crucial role in the development of outgroup prejudice. For example, if identity development occurs in circumstances of high intergroup conflict, social categorization, identity development, and intergroup biases are accelerated (Bar Tal & Teichman, 2005) as compared to situations of low conflict. Also, individual experiences with intergroup contact might play a crucial role here that remains to be further explored.

Clearly, our identity development related explanation of age differences in prejudice-related affective priming effects, remains rather tentative, as we did not directly access the identity status of our participants which remains to be explored by future research. We are also aware that the cross-sectional designs of our studies limit the generalizability of results for developmental processes in prejudice automatization. However, given that the three studies with German samples were conducted during the course of 3 years (implying that participants of 8<sup>th</sup> and 9<sup>th</sup> grade in Study 3 attended 5<sup>th</sup> and 6<sup>th</sup> grade when Study 1 was conducted), we deem it rather unlikely that cohort effects could explain the age differences. Of course, it is desirable to investigate prejudice automatization during childhood and adolescence in longitudinal designs. This would further allow exploration of antecedents and conditions of developmental changes observed in the current studies. In this vein the relations between automatic prejudice activation and open expression of prejudice as well as discriminatory

behaviors need to be further researched. In the current studies, open prejudice expression and prejudice-related affective priming scores were unrelated. Also, older children did not express higher (but rather a lower) levels of prejudice as compared to younger children. Rather than suggesting a dissociation between attitude activation and attitude expression, we point out that (a) there was a rather low conceptual overlap between the measures with the priming procedure tapping rather affective representations and the questionnaires tapping rather cognitive representations and that (b) the affective priming was characterized by low internal consistencies limiting the application as an interindividual difference measure (though they allowed to assess group differences). Thus, so far, conclusions about dissociations between automatic prejudice activation and open prejudice expression in children and adolescents should be drawn with caution and future research is needed to fully understand this relation.

## References

- About, F. (1988). *Children and prejudice*. Oxford, GB: Blackwell.
- About, F. E. (2008). A social-cognitive developmental theory of prejudice. In S. E. Quintana & C. McKown (Eds), *Handbook of race, racism and the developing child*. (pp. 55-71). Hoboken, NJ: Wiley
- Balke, D., El-Menouar, Y., & Rastetter, R.U. (2002). Einschätzung von Kosten durch Zuwanderer. [Assessment of costs, caused by immigrants] *Zentralarchiv für Empirische Sozialforschung* (1998). ISSP 1995 - National identity. Codebook ZA-No. 2880, Cologne.
- Banaji, M. R., Baron, A. S., Dunham, Y., & Olson, K. (2008). The development of Intergroup social cognition: Early emergence, implicit nature, and sensitivity to group status. In S. R. Levy & M. Killen (Eds.), *Intergroup attitudes and relations in childhood through adulthood*. New York, NJ: Oxford.
- Baron, A. S. & Banaji, M. R. (2006). The development of implicit attitudes. Evidence of race evaluation from ages 6 and 10 and adulthood. *Psychological Science*, 7, 53-58.
- Bar-Tal, D. & Teichman, y. (2005). *Stereotypes and prejudice in conflict, Representations of Arabs in Israeli Jewish society*. Cambridge University Press.
- Bennett, M., & Sani, F. (2003). The role of target gender and race in children's encoding of category-neutral person information. *British Journal of Developmental Psychology*, 21, 99–112.

Bigler, R. S., & Liben, L. S. (2006). A developmental intergroup theory of social stereotypes and prejudice. In R. V. Kail (Ed.), *Advances in Child Development and Behavior*, (Vol. 34, pp. 39–89). San Diego: Elsevier.

Boehnke, K., Hagen, J., & Hefner, G. (1998). On the development of xenophobia in Germany: The adolescent years. *Journal of Social Issues* 54, 585-602.

Bosson, J.K., Swann, W.B. Jr., & Pennebaker, J.W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology*, 79, 631–643.

Cunningham, W.A., Preacher, K.J., & Banaji, M.R. (2001). Implicit attitude measures: Consistency, stability, and convergent validity. *Psychological Science*, 12, 163–170.

De Houwer, J. (2001). A structural and process analysis of the Implicit Association Test. *Journal of Experimental Social Psychology*, 37, 443-451.

De Houwer, J. (2003). 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: Erlbaum.

Brendl, M., Markman, A., & Messner, C. (2001). How do indirect measures of evaluation work? Evaluating the inference of prejudice in the Implicit Association Test. *Journal of Personality and Social Psychology*, 81, 760-773.

Centraal Bureau voor de Statistiek Nederlands (2009). Den Haag/Heerlen.

Chang, B., Mitchell, C.J. (2009). Processing fluency as a source of salience asymmetries in the Implicit Association Test. *Quarterly Journal of Experimental Psychology*, 26, 1-25.

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology*, 43, 972-978.

De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit measures: A normative analysis and review. *Psychological Bulletin*, 135, 347–368.

Degner, J. (2009). On the (Un)Controllability of affective priming: Strategic manipulation is feasible but can possibly be prevented. *Cognition and Emotion*, 23 , 327-354

Degner, J. & Wentura, D. (2009). Not everybody likes the thin and despises the fat: One's weight matters in the automatic activation of weight-related social evaluations. *Social Cognition*, 27, 202-220.

Degner, J., Wentura, D., Gniewosz, B., & Noack, P. (2007). Hostility-related prejudice against Turks in adolescents: Masked affective priming allows for a differentiation of automatic prejudice. *Basic and Applied Social Psychology*, 29, 245-256.

Degner, J., Wentura, D., & Rothermund, K. (2006). Indirect assessment of attitudes with response-time-based measures. Chances and problems. *Social Psychology (former Zeitschrift für Sozialpsychologie)*, 37, 131-139.

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

Draine, S. C., & Greenwald, A. G., (1998) Replicable unconscious semantic priming. *Journal of Experimental Psychology: General*, *127*, pp. 286-303.

Dunham, Y., Baron, A.S., & Banaji, M.R. (2006). From American city to Japanese village: A cross-cultural investigation of implicit race attitudes. *Child Development*, *77*, 1268-1281.

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.

Fazio, R. H., Sanbonmatsu, D. M., Powell, M. C., & Kardes, F. R. (1986). On the automatic activation of attitudes. *Journal of Personality and Social Psychology*, *50*, 229-238.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297-327.

Fiske, S. T. (2005). Social cognition and the normality of prejudgment. In J. F. Dovidio, P. Glick, & L. A. Rudman (Eds.), *On the nature of prejudice: Fifty years after Allport*. (pp. 36-53). Malden, MA: Blackwell.

French, S. E., Seidman, E., Allen, L.R., & Aber, J. L. (2006). The development of ethnic identity during adolescence. *Developmental Psychology*, *42*, 1-10.

Frindte, W., Funke, F., & Waldzus, S. (1996). Xenophobia and right-wing-extremism in German youth groups--some evidence against unidimensional misinterpretations. *International Journal of Intercultural Relations, 20*, 463-478.

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (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., & 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.

Han, H. A., Olson, M. A., & Fazio, R. H. (2006). The influence of experimentally created extrapersonal associations on the Implicit Association Test. *Journal of Experimental Social Psychology, 42*, 259-272.

Hirschfeld, L. A. (2008). Children's developing conceptions of race. In S. E. Quintana & C. McKown (Eds), *Handbook of race, racism and the developing child*. (pp.37-54). Hoboken, NJ, Wiley.

Klauer, K. C., & Musch, J. (2003). Affective priming: Findings and theories. In J. Musch, & K. C. Klauer (Eds.), *Psychology of Evaluation: Affective Processes in Cognition and Emotion* (pp. 7-49). Mahwah, NJ: Erlbaum.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the „Who Said What?“ paradigm. *Journal of Experimental Social Psychology, 75*, 1155-1178.

Lang, P.J., Bradley, M.M., & Cuthbert, B.N. (2005). International affective picture system (IAPS): Affective ratings of pictures and instruction manual. Technical Report A-6. University of Florida, Gainesville, FL.

Livingston, R. W. & Brewer, M. B. (2002). What are we really priming? Cue-based versus category-based processing of facial stimuli. *Journal of Personality and Social Psychology, 82*, 5-18.

Mackie, D. M., & Smith, E. R. (1998). Intergroup relations. Insights from a theoretically integrative approach. *Psychological Review, 105*, 499-529.

Macrae, C. N., Quinn, K. A., Mason, M. F., & Quadflieg, S. (2005). Understanding others: The face and person construal. *Journal of Personality and Social Psychology, 89*, 686-695.

Masson, C.N. & Verkuyten, M.J.A.M. (1993). Prejudice, ethnic identity, contact and ethnic group preferences among young Dutch adolescents. *Journal of Applied Social Psychology, 23*, 156-168.

Nesdale, D. (1999). Social identity and ethnic prejudice in children. In P. Martin & W. Noble (Eds.). *Psychology and society*. (pp 92-110). Brisbane: Australian Academic Press.

Nesdale, D. (2001). Development of prejudice in children. In M. Augustinos & K. Reynolds (Eds.). *Understanding prejudice, racism, and social conflict*. (pp 57-72). London: Sage.

Newman, B. M. & Newman, P. R. (2002). Group identity and alienation: Giving the we its due. *Journal of Youth and Adolescence*, 30, 515-538.

Olson, M. A., & Fazio, R. H. (2004). Reducing the influence of extrapersonal associations on the Implicit Association Test: Personalizing the IAT. *Journal of Personality and Social Psychology*, 86, 653-667.

Olson, M. A., & Fazio, R. H. (2003). Relations between implicit measures of prejudice: What are we measuring? *Psychological Science*, 14, 636-640.

Otten, S. & Wentura, D. (1999). About the impact of automaticity in the Minimal Group Paradigm: Evidence from affective priming tasks. *European Journal of Social Psychology*, 29, 1049-1071.

Pettigrew, T. F., & Meertens, R. W. (1995). Subtle and blatant prejudice in Western Europe. *European Journal of Social Psychology*, 25, 57-75.

Phinney, J. S. & Chavira, V. (1992). Ethnic identity and self-esteem: an exploratory longitudinal study. *Journal of Adolescence*, 15, 271-281.

Pollack, I. (1970). A nonparametric procedure for evaluation of true and false positives. *Behaviour Research Methods and Instrumentation*, 2, 155-156.

Quintana, S. M. (1998). Children's developmental understanding of ethnicity and race. *Applied and Preventive Psychology*, 7, 27-45.

Rothermund, K., & Wentura, D. (2004). Underlying processes in the Implicit Association Test (IAT): Dissociating salience from associations. *Journal of Experimental Psychology: General*, *133*, 139-165.

Rudman, L. A. (2004). Sources of implicit attitudes. *Current Directions in Psychological Science*, *13*, 79-82.

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development*, *76*, 451-466.

Scherer, L. D. & Lambert, A. J. (2009). Contrast Effects in Priming Paradigms: Implications for Theory and Research on Implicit Attitudes. *Journal of Personality and Social Psychology*, *97*, 383–403.

Stangor, C., Lynch, L., Duan, C., & Glass, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*, 207-218.

Taylor, S.E., Fiske, S. T., Etoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology*, *36*, 778-793.

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Wagner, U., van Dick, R., Pettigrew, T. F., & Christ, O. (2003). Ethnic prejudice in East and West Germany: The explanatory power of intergroup contact. *Group Processes and Intergroup Relations*, *6*, 22-36.

Wentura, D. & Degner, J. (in press). A practical guide to sequential Priming and related tasks. Prepared for In B. Gawronski & B. K. Payne (Eds.), *Handbook of Implicit Social Cognition: Measurement, Theory, and Applications*. New York, NY: Guilford Press.

Wentura, D., Kulfanek, M., & Greve, W. (2005). Masked affective priming by name letters: Evidence for a correspondence of explicit and implicit self-esteem. *Journal of Experimental Social Psychology, 41*, 654-663.

Woellert, F., Kröhnert, S., Sippel, L., & Klingholz, R. (2009). Ugenutzte Potentiale: Zur Lage der Integration in Deutschland. [Unused potential. Integration in Germany]. Berlin-Institut fuer Bevoelkerung und Entwicklung. Berlin.

## Footnotes

<sup>1</sup> Cronbach's Alpha refers to the reliability of the picture scores by taking the raters as "items".

<sup>2</sup> Because of different settings of computer screens, timing of stimulus presentation was not exactly the same for all participants. Specifically, the parameters described above applied to the majority of 198 participants (60 Hz screens). For another 86 participants, the primes were presented for 253 ms (75 Hz screens; SOA = 470 ms, response deadline = 1400 ms). For another 20 participants, prime presentation lasted for 365ms (52 Hz screens; SOA = 540 ms, response deadline = 2020 ms). However, timing of stimulus presentation did not affect results.

<sup>3</sup> For all trend analyses throughout the paper, quadratic and cubic trends were non-significant and are therefore not reported in detail.

<sup>4</sup> Of course, these analyses have to be interpreted with caution. If, for example, both indices would have yielded a comparable age trend, the most straightforward interpretation would have been to assume an age trend in the difference between responses to negative versus positive target stimuli, irrespective of priming effects.

<sup>5</sup> Contrary to Study 1, a significant priming effect was found in the error rates for standard positive and negative primes ( $M = 2.34$ ,  $SD = 5.94$ ,  $t(140) = 4.68$ ,  $p < .001$ ,  $d = .39$ ), but not for the prejudice related primes ( $M = 0.68$ ,  $SD = 4.71$ ). However, priming effects in error scores were unrelated to age,  $F_s < 1$ , *ns*.

<sup>6</sup> Applying this paradigm to children, Bennett and Sani (2003) found indeed significantly more within- than between-race confusions, with no age differences between 5-, 8-, and 11-year old children, which contradicts our hypothesis. However, the age-range in this study does not include young adolescents and it is difficult to draw conclusions about higher age groups. Furthermore, this study involved stimulus targets who only differed according to one category (i.e., ethnicity), limiting the generalizability of results to naturally occurring contexts in which people vary in terms of multiple categorizations (e.g., age, ethnicity, gender).

<sup>7</sup> This practice phase was included for children of both conditions to counter an alternative explanation of age differences in the affective priming task in Study 1 to 3 and differences between the affective priming score and the IAT in Study 3 proposed by a reviewer of an earlier version of this manuscript. These differences could be explained by assuming that younger children simply lack the ability to correctly classify the stimulus pictures according to their ethnicity and show prejudice effects in the IAT only because of a practice phase in which they are familiarized with a hitherto unknown classification of Turks and Germans and explicitly trained to apply it to the target pictures. If this alternative explanation would apply, no differences between age groups and conditions should occur in Study 4 because children in both conditions went through the same practice phase.

<sup>8</sup> We tested whether spontaneous use of ethnicity in the Who-said-What paradigm mediated the age effect on affective priming in the *no categorization* condition. The affective priming score as well as the ethnicity bias in the Who-said-What paradigm were significantly related to participants age with  $\beta = .34$ ,  $t(62) = 2.88$ ,  $p = .005$  and  $\beta = .27$ ,  $t(62) = 2.20$ ,  $p = .03$ , respectively. More importantly, category use in the Who-said-what paradigm significantly predicted affective priming scores if age was controlled for,  $\beta = .27$ ,  $t(62) = 2.20$ ,  $p = .03$ . The mediation of the age trend in priming by category use was significant as indicated by a Sobel test,  $z = 2.20$ ,  $p = .03$ . However, mediation was only partial and the effect of age on affective priming scores remained significant in the multiple regression with age and category use as predictors,  $\beta = .24$ ,  $t(62) = 2.01$ ,  $p = .05$ . The opposite mediation pattern (i.e. automatic evaluations in the affective priming task mediating the relation between participants' age and their category use in the Who-said-What paradigm) was not significant,  $z = 1.57$ ,  $p = .12$ . However, because of the low reliability of the affective priming score and possible confounds due to the fixed order of the tasks, results of the mediation analysis have to be interpreted with caution.



Table 1

*Means (Standard deviations in parentheses) of explicit prejudice measures in Study 1 to 3.*

		alpha	5th	6th	7th	8th	9th
Study 1	Blatant prejudice	.74	3.17 (1.11)	3.04 (1.04)	2.72 (0.96)	2.66 (1.04)	
	General Xenophobia	.83	3.24 (1.02)	3.46 (0.94)	3.19 (1.04)	3.17 (0.86)	
Study 2	Blatant prejudice	.78	3.07 (0.67)	2.99 (0.51)	3.00 (0.44)	2.98 (0.42)	
	General Xenophobia	.74	3.22 (0.92)	3.23 (0.94)	3.09 (1.04)	2.98 (0.91)	
	Rating Turks	.48	3.31 (1.14)	3.67 (1.31)	3.56 (1.03)	3.85 (1.07)	
	Rating Germans	.78	4.73 (0.88)	4.67 (1.17)	4.74 (0.79)	4.76 (0.69)	
Study 3	Rating Turks	.92	3.26 (1.46)	3.39 (1.43)	3.78 (1.19)	3.09 (1.04)	3.32 (1.15)
	Rating Germans	.81	4.48 (0.99)	4.64 (0.78)	4.63 (0.68)	4.74 (0.91)	4.47 (0.90)

Table 2

*Percentages of correct and erroneous person assignments (Standard deviations in parentheses) of correctly as old remembered statements in the Who-Said-What Paradigm in Study 4.*

	within ethnicity				between ethnicity	
	correct	within age*		between age	within age	between age
5 <sup>th</sup> (n = 27)	46.14 (17.81)	31.87 (14.78)	[15.93 (7.39)]	8.14 (8.20)	21.22 (11.46)	8.56 (5.92)
6 <sup>th</sup> (n = 30)	47.00 (20.39)	34.71 (21.47)	[17.36 (10.74)]	7.63 (6.31)	19.25 (10.98)	8.77 (7.54)
7 <sup>th</sup> (n = 34)	49.58 (15.09)	36.74 (17.28)	[18.37 (8.64)]	11.21 (8.59)	14.57 (9.83)	6.26 (5.91)
8 <sup>th</sup> (n = 32)	64.03 (13.60)	32.82 (17.67)	[16.41 (8.83)]	5.71 (4.67)	9.55 (7.44)	4.29 (5.00)

\* unadjusted values in square brackets

## Figure captions

*Figure 1*

Age effects on the prejudice-related priming effects in Study 1 (error bars represent standard errors).

*Figure 2*

Age effects on the prejudice-related priming effects in Study 2 (error bars represent standard errors). Note that the dependent variable is the error rate due to using the response-window procedure.

*Figure 3*

Age effects on the prejudice-related priming effects in Study 3 (error bars represent standard errors).

*Figure 4*

Age effects on the prejudice-related priming effects in Study 4 (error bars represent standard errors).

Figure 1

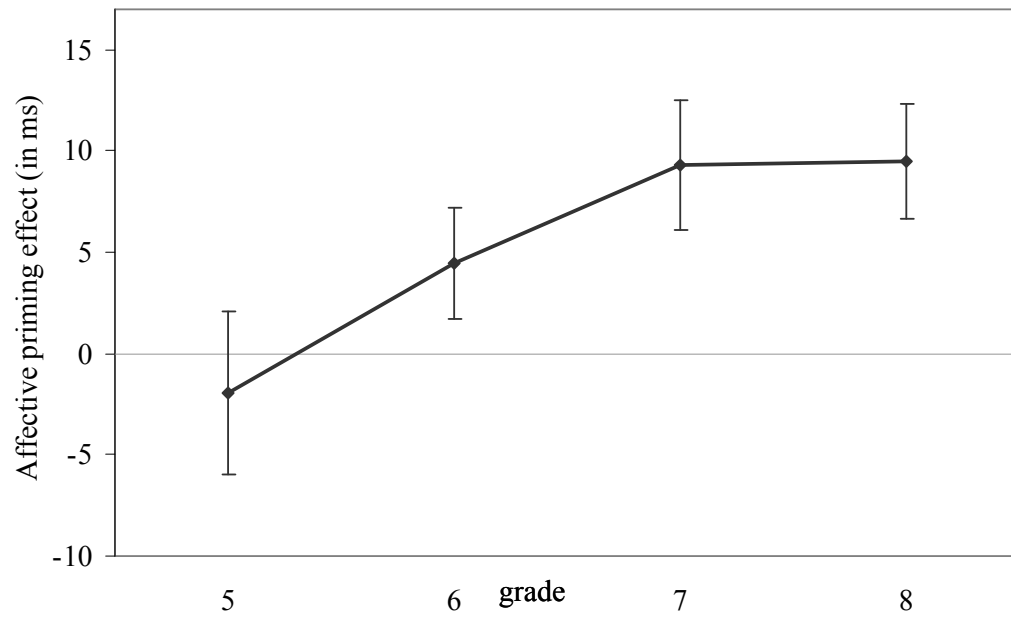


Figure 2

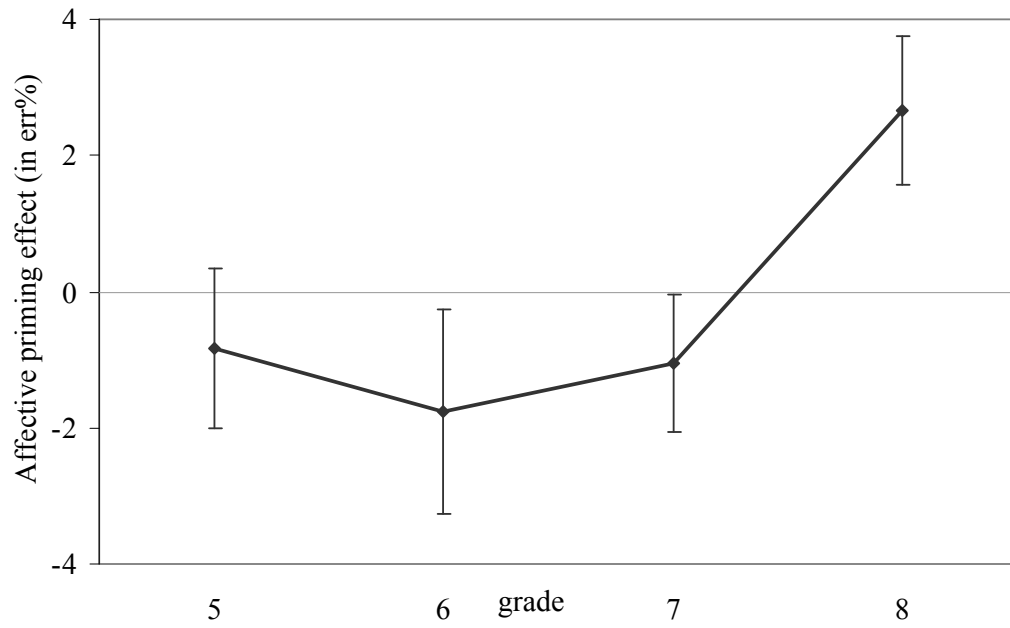


Figure 3

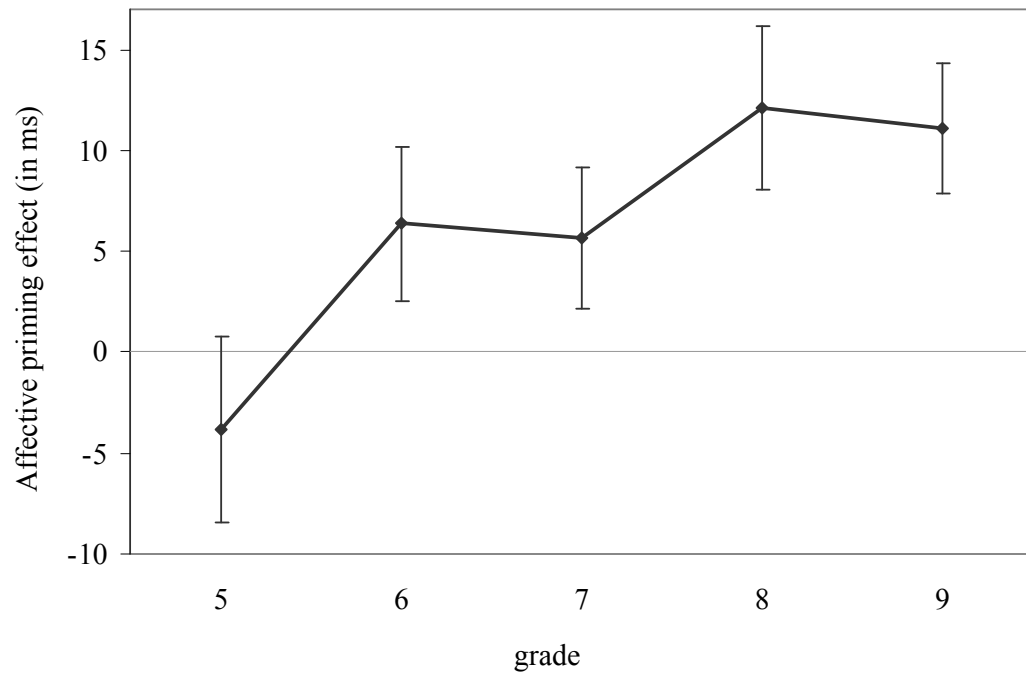


Figure 4

