The Effectiveness of Associative and Rational Statistical Learning in Reducing Children's Stereotype Formation

by

Vera Bingchen Chai

Hon. B. Sc., University of Toronto, 2012

A thesis presented to Ryerson University

in partial fulfillment of the

requirements for the degree of

Master of Arts

in the program of

Psychology (Psychological Science)

Toronto, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I authorize Ryerson University to lend this thesis to other institutions or individuals for the purpose of scholarly research

I further authorize Ryerson University to reproduce this thesis by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

I understand that my thesis may be made electronically available to the public.

Abstract

The Effectiveness of Associative and Rational Statistical Learning in Reducing Children's Stereotype Formation

Master of Arts, 2015

Vera Bingchen Chai

Psychology (Psychological Science)

Ryerson University

A stereotype is a rigid and overgeneralized belief about the characteristics of a social group. Stereotyping is a pervasive phenomenon, and has detrimental effects on children's development such that it leads to biased information processing and stereotype threat. One of the underlying mechanisms for stereotype formation is illusory correlation, which refers to the erroneous inference about the relationship between two categories of events that in fact are uncorrelated. Given that most of the stereotype reduction training is focused on adults rather than children, this Master's thesis aimed to examine the effectiveness of two methods that could potentially reduce stereotyping in children. More specifically, this work investigated whether facilitating associative and rational statistical learning could reduce stereotyping in children through inhibiting the formation of illusory correlation. The results showed that 5- to 10-year-old children consistently perceive an illusory correlation between the numerically smaller minority group and the infrequently occurring, negative behaviour. However, the perception of an illusory correlation among 5- to 8-year-olds was significantly reduced through the facilitation of statistical learning, but not associative learning.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Dr. Lili Ma for the continuous support of my Master's study and research, for her caring, patience, motivation, and immense knowledge. Her guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor and mentor for my Master's study.

Besides my advisor, I would like to thank the rest of my thesis committee: Dr. Becky Choma, and Dr. Julia Spaniol, for their encouragement and insightful comments. My sincere thanks also goes to Dr. Lixia Yang for enlightening me the first glance of research.

I thank my labmates in the Early Childhood Cognition Lab for their assistance in recruitment and data collection, for the stimulating discussions, and for all the fun we have had in the last two years.

Last but not least, I would like to thank my family: my parents Yanrong Gao and Yulong Chai, and my husband, Genan Yang. They were always there supporting me and stood by me through the good times and bad.

Table of Contents

List of Tables

List of Figures

List of Appendices

Introduction

A stereotype is a rigid and overgeneralized belief about the characteristics of a social group (Allport, 1979; Devine, 1989). Although stereotypes link people to characteristics that are negative, neutral, or positive, they may all lead to negative consequences as they give rise to biased or unfair expectations (Biernat & Vescio, 2002; McCabe & Brannon, 2004). Studies have shown that children as young as 3 to 4 years demonstrate awareness and endorsement of different types of stereotypes (e.g., racial and gender stereotypes; Aboud, 1988; Miller, Lurye, Zosuls, & Ruble, 2009). Social psychologists have suggested that stereotyping has adaptive values as it can facilitate information processing (e.g., Macrae, Hewstone, & Griffiths, 1993). Yet, stereotyping has harmful effects on children because it can lead to biases in information processing (Bigler & Liben, 1992, 1993) and induce stereotype threat (Kit, Tuokko, & Mateer, 2008). Social psychologists have proposed various approaches to explaining the formation of stereotypes among children. One well-studied mechanism is the formation of illusory correlation, which is the erroneous inference about the relation between two categories of events (Hamilton & Gifford, 1976). Despite these findings, past research and interventions have been focused mostly on stereotype reduction in adults. Little attention has been paid to how to reduce stereotypes in children.

It has been shown that stereotyping in adults can be reduced through associative learning (Murphy, Schmeer, Vallee-Tourangeau, Mondragon, & Hilton, 2011). Another potential method for stereotype reduction is the promotion of critical evaluation of a stereotype's validity, which can be done by facilitating one's ability to make rational statistical inference. Rational statistical inference refers to the inductive inference mechanism that takes into account statistical information such as sample size, variability, and validity (Xu, 2007). Studies have shown that children are able to make rational statistical inference from a young age. For example, 15-month-olds consider the characteristics of a sample (e.g., random or nonrandom) when making generalizations from the sample to the population (Gweon, Tenenbaum, & Schulz, 2010). In addition, 2-year-olds are capable of revising their prior beliefs after observing statistical sampling evidence that is inconsistent with their own beliefs (Ma & Xu, 2011).

The overarching goal of this Master's thesis is to examine the effectiveness of associative versus rational statistical learning in reducing children's stereotype formation. The specific objectives of this research are to examine: 1) whether children between the ages of 5 and 10 form stereotypes on the basis of illusory correlation, 2) whether increased associative learning can reduce the formation of illusory correlation among children, and 3) whether the training of rational statistical learning can reduce the formation of illusory correlation among children.

Literature Review

Stereotype as energy-saving mental device. As noted earlier, stereotype is a rigid and overgeneralized belief about the characteristics of a social group (Allport, 1979; Devine, 1989). According to Devine (1989), the development and activation of stereotypes is inevitable as it is an automatic process and can only be inhibited by intentionally replacing stereotypes with non-stereotypical beliefs. One possible explanation for the persistence and prevalence of stereotyping is that stereotypes have adaptive values that might facilitate cognitive functions by saving cognitive resources. Social psychologists argue that stereotypes may function as resource-preserving mental devices that simplify information processing and response generation. Specifically, given that reality is too complicated for any person to represent accurately, stereotyping can help preserve cognitive resources by directing attention toward information that is most easily understood and away from the information that is most difficult to encode (e.g., Bodenhausen & Lichtenstein, 1987; Stangor & Duan, 1991; Taylor & Crocker, 1981). For example, Fiske and Neuberg (1990, p. 14) have remarked that "we are exposed to so much information that we must in some manner simplify our social environment… for reasons of cognitive economy, we categorize others as members of particular groups – groups about which we often have a great deal of generalized, or stereotypic, knowledge."

Supporting evidence for this notion has demonstrated that individuals tend to rely on stereotypes to a greater extent when they lack the ability or motivation to think more deeply about members of stereotyped groups. Situational factors that affect the degree of stereotyping include physical depletion (Bodenhausen, 1990; Kim & Baron, 1988), task difficulty (Bodenhausen & Lichtenstein, 1987; Pratto & Bargh, 1991), multiple task demands (Gilbert & Hixon, 1991; Macrae et al., 1993), anxiety-induced arousal (e.g., Wilder & Shapiro, 1989), and positive moods (Stroessner & Mackie, 1993). For instance, in a study done by Macrae, Milne, and Bodenhausen (1994), participants were asked to learn different traits about different individuals and form impressions about these individuals while listening to a passage about Indonesia. Half of the participants were assigned to the stereotype-consistent condition in which they learned traits that are consistent with the stereotype towards those individuals. The other half of the participants was assigned to the neutral condition in which the traits they learned were inconsistent with the stereotype. Results of this study showed that participants in the stereotype-consistent condition not only remembered more traits but also scored higher on the multiple-choice test of their knowledge about Indonesians than participants in the neutral condition.

In summary, the above-mentioned researchers posit that stereotyping persists because it functions as a "mental tool" that allows stereotype perceivers to free up limited cognitive resources for the execution of other mental activities (Fiske, 1989; Gilbert & Hixon, 1991; Macrae et al., 1994; Sherman & Corty, 1984).

The effects of stereotyping on development. Despite the findings that stereotyping has benefits of preserving cognitive resources, research in developmental psychology suggests that the costs of stereotyping are unacceptable and significantly outweigh its benefits as it may have adverse effects on children's development.

First, stereotyping has a detrimental impact on individuals who stereotype because it leads to biased information processing. Past research has shown that when processing information about people, children tend to make errors that are consistent with their own stereotypic beliefs, such that they distort information based on gender, racial, or age stereotypes (e.g., Bigler & Liben, 1992, 1993; Davidson, Carmeron & Jergovic, 1995). For example, examining the effect of "beauty is good" stereotype on children's information processing, Ramsey and Langlois (2002) found that stereotypic processing causes children to bias information about both adult and child female targets based on the level of attractiveness. In this study, an experimenter read 8 age-appropriate stories to 3- to 7-year-old children. Each story depicted a child narrator encountering 2 target characters of the same gender. Among these 8 stories, 4 depicted male characters and 4 depicted female characters. During the encounter, one of the target characters displayed a positive trait, and the other one displayed a negative trait. Children were then presented with the photographs of these two characters. One of the characters was attractive, and the other one was unattractive. Within 5 to 10 minutes of finishing the story, children were tested about the contents of the stories. Results showed that children were more likely to identify female characters when their traits were consistent with the "beauty is good" stereotype. In addition, children were more likely to make mistakes and choose attractive females as displaying positive traits when in fact the unattractive females had displayed the positive traits (Ramsey & Langlois, 2002).

Studies investigating the effect of negative stereotypes about individuals from a disadvantaged social class have obtained similar findings. For example, in a study involving 153 elementary school students, Désert and colleagues (2009) found that as early as in the first grade, children show biased information processing towards people with lower socioeconomic status. These children, regardless of their own socioeconomic status, considered students from a higher social class as more likely to succeed in school than children from a lower socioeconomic class (Désert Préaux, & Jund, 2009).

Second, stereotyping has harmful effects on targets of stereotypes by fostering stereotype threat and subsequently undermining children’s school performance. Stereotype threat is the situational phenomenon in which a member of a stigmatized group experiences anxiety that he or she might potentially confirm a negative stereotype about his or her social group (Steele & Aronson, 1995). It was first demonstrated by Steele and Aronson (1995). They found that African American adults underperformed in relation to Caucasian adults in difficult verbal tests when a negative domain-relevant in-group stereotype was made salient. However, when the in-group stereotype was not salient, African American adults’ test performance was not significantly different from that of Caucasian adults. Developmental researchers later conducted similar studies in children and found that stereotype threat can also undermine children’s performance in various task domains, such as mathematics (Tomasetto, Alparone, & Cadinu, 2011), spatial memory (Neuburger, Jansen, Heil, & Quaiser-Pohl, 2012), and logic reasoning (Désert et al., 2009). According to Schmader, Johns and Frobes (2008), such impaired performance is due to the fact that stereotype threat triggers three different mechanisms that further reduce children's working memory capacity, including physiological stress, monitoring processes, and emotion-suppressing efforts.

In a study involving 476 elementary school boys and girls, Muzzatti and Agnoli (2007) investigated the impact of gender stereotypes on school children's mathematics performance. In this study, children were randomly assigned to the experimental or the control condition. In the experimental condition, the mathematics gender stereotype that men are better at mathematics than women was made salient. Specifically, they presented a vignette portraying 10 famous mathematicians, including 9 men and 1 woman. In the control condition, the mathematics gender stereotype was not made salient and the vignette portrayed 10 neutral stimuli, including 9 flowers and 1 fruit. Moreover, children in the experimental and control conditions were interviewed regarding their opinions of mathematics and tested on their mathematical ability. The results showed that starting from the fourth grade, girls' math performance was significantly impaired in the experimental condition than in the control condition. Additionally, compared to the control condition, fourth and fifth grade girls in the experimental condition reported significantly lower self-confidence in mathematics, perceived mathematics as less likeable, and rated boys as better in mathematics than girls. Interestingly, Muzzatti and Agnoli (2007) found that the age when children started to endorse the mathematics gender stereotype coincided with the age children started to demonstrate a gender gap in math performance. In other words, school children do not show significant gender differences in math performance until they start endorsing the gender stereotype that males are better at mathematics than females.

Taken together, these findings demonstrate that stereotyping has detrimental impacts on the children who stereotype and the stereotyped targets. For children who stereotype, stereotyping can lead to biased information processing. As for the stereotyped targets, stereotyping not only has short-term effects on children's performance under certain situations, but also has potentially long-term impacts on children's skill acquisition in stereotype-related domains by influencing their beliefs and attitudes towards those domains.

Stereotype formation. Given these detrimental impacts of stereotypes on children's development, several developmental approaches have been proposed to explain the formation of stereotypes among children. A widely studied approach is the social learning approach (Bandura, 1986). According to this approach, children acquire knowledge about stereotypes from the social environment in which they live. The primary source of the stereotype knowledge includes parents and family members who provide information and reinforce stereotype contents. Other sources that also play an influential role in children's acquisition of stereotypes are the media, peer groups, and schools. Furthermore, according to this approach, stereotypes can also be learned on the basis of children's real life observations of differences between groups (e.g., Eagly & Steffen, 1984; Goodman, 1952; Liebert, Sobol, & Copemann, 1972).

According to the psychodynamic approach, stereotype formation is the result of childhood emotional experiences. Negative stereotypes are thought to reflect children's intrapersonal conflict or maladjustment, which develops as a consequence of harsh and punitive parenting style (e.g., Adorno, Frenkel-Brunswick, Levinson, & Sanford, 1950; Allport & Kramer, 1946; Bettelheim & Janowitz, 1950).

Lastly, the cognitive developmental approach emphasizes the qualitative changes occurring in children's cognitive development as a basis for stereotype acquisition (Aboud, 1988; Katz, 1976). According to this approach, one of the cognitive abilities that might be related to stereotype formation is the ability to categorize (Piaget, 1965). Although preschoolers are able to sort objects along a single dimension, they have difficulty in understanding that the same object or person could simultaneously belong to two categories (Liben & Bigler, 1987). This inability in sorting objects or individuals along multiple dimensions leads to difficulties in processing counter-stereotypic information, and hence, forms a basis for stereotype development.

One intriguing and well-studied mechanism for stereotype development is the formation of illusory correlation (Hamilton & Sherman, 1989). Chapman (1967) defined the illusory correlation as the erroneous inference about the relation between two categories of events that in fact are uncorrelated or are correlated to a lesser degree than the perceived correlation. In the classic illusory correlation paradigm, Hamilton and Gifford (1976) demonstrated that adults tend to overestimate the frequency of co-occurrence of two distinct events, such that they tend to overestimate how often members from a minority group (i.e., infrequently occurring group) perform an infrequent undesirable behaviour (i.e., infrequently occurring behaviour). In their study, participants were asked to read statements about positive (and more frequent) and negative (less frequent) behaviours of a majority group and a minority group. The ratio of positive to negative behaviours was the same for these two groups, so there was no actual relation between group membership and behaviour. Participants were then asked to estimate the frequency of occurrence for positive and negative behaviours in each group and complete evaluative ratings of these two groups. Results showed that participants could accurately estimate the frequency of occurrence for behaviours committed by the majority group. However, they overestimated the frequency of occurrence for the less frequent behaviours (in this case, the negative behaviours) committed by the minority group. In addition, participants rated the majority group more positively than the minority group. Hence, illusory correlation can, at least partially, explain the formation of stereotypes as participants perceive two groups of stimulus individuals as being different from each other when no informational basis for such difference was available (Hamilton & Gifford, 1976).

Studies investigating the development of stereotypes among children have demonstrated that children's patterns of social perceptions are consistent with those findings of adults' illusory correlation. Primi and Agnoli (2002), for example, investigated children's perceptions of a novel social group by using a paradigm similar to the one used in Hamilton and Gifford (1976). Primi and Agnoli (2002) found that just like adults, first-, third-, fifth-, and seventh-grade children readily perceived illusory correlation, that is, they associated members from a minority group with negative behaviours even though negative behaviours occurred at a lower frequency for the minority group. In a follow-up study, Johnston and Jacobs (2003) investigated illusory correlation among second- and fifth-grade children using a similar methodology except that the frequency of positive and negative behaviours was reversed (in this case, positive behaviours occurred at a lower frequency than negative behaviours). The results showed that after reversing the frequency, children perceived an illusory correlation between the minority group and positive behaviours and rated the minority group more positively than the majority group. Therefore, these findings suggest that children, just like adults, perceive illusory correlation that makes them associate minority group with infrequently occurring behaviours.

To explain the phenomenon of illusory correlation, researchers proposed a paradigm called the distinctiveness-based explanation (Hamilton & Gifford, 1976; Hastie & Park, 1986). This distinctiveness-based explanation asserts that during the presentation of information, infrequent items are more distinctive than frequent items. The infrequent items further receive more enhanced encoding and thus become more accessible when judgments are made. To test this theory, Hamilton, Dugan and Trolier (1985) conducted an experiment using the same methodology as the one used in Hamilton and Gifford's (1976) study, except that participants were divided into three conditions. In the first condition, participants were presented with behaviour items using the traditional illusory correlation paradigm and demonstrated the expected illusory correlation effect. In the second condition, without having read the stimulus sentences, participants were presented only with a frequency summary table revealing the distribution of desirable and undesirable behaviours performed by the majority and the minority group. They were instructed to imagine that they had read 39 sentences that were distributed as shown in the table. Participants in the second condition did not demonstrate the perception of illusory correlation. Lastly, in the third condition, participants were presented with the behaviour items and the frequency summary table. By showing both the behaviour items and the frequency summary table, it was expected that participants would be less likely to form illusory correlation. Yet, participants in this condition still reported liking the majority group more than the minority group, even though they had access to the summary table. The results of this experiment demonstrated that the formation of illusory correlation was due to biased encoding of sequentially presented information. The results contradicted with the explanation that participants gave more favourable ratings to the majority group because the absolute difference between the frequency of desirable and undesirable behaviours was greater for the majority group than for the minority group. If this were the case, participants in the summary-table-only condition would rate the majority group as more likeable as the absolute differences were maximally obvious in this condition. However, participants in this condition did not differentially evaluate the majority and the minority group. Given the evidence that individuals were able to make accurate frequency estimations and judgments toward the minority group when merely given a summary table, Hamilton et al. (1985) suggested that the enhanced accessibility of the distinctive information prevented individuals from critically evaluating the validity of their current belief by using the incoming statistical evidence.

In summary, these findings suggest that illusory correlation could in part explain the formation of stereotypes toward minorities or infrequently encountered groups. Specifically, a stereotype is formed due to the distinctiveness of statistically infrequent information, such that the distinctiveness gives rise to biased encoding and recalling of information and further prevents people from making rational statistical inferences. This type of cognitive bias does not appear to diminish as children grow older. Rather, it seems to persist even after they enter adulthood. Thus, to prevent the development of stereotypes, stereotype reduction interventions should be employed at an early stage during childhood.

Stereotype reduction through training. Despite the compelling evidence that stereotype formation emerges early in development and has significant harmful effects on children's development, past research on stereotype reduction and intervention has predominantly targeted adults (Dovidio et al., 2004; Galinsky & Moskowitz, 2000; Stewart & Payne, 2010). Researchers have developed cognitive interventions, for example, that aim to reduce stereotype formation by promoting rational statistical reasoning process (i.e. Schaller, Asp, Rosell, & Heim, 1996; Schaller & O'Brien, 1992).

Schaller et al. (1996) demonstrated that training the logic of analysis of covariance (ANCOVA) can significantly increase the accuracy of group impressions that individuals form. During the training session, participants were taught to emphasize on drawing inferences about covariation and focus on three-dimensional statistical thinking. As an example, in the training, participants were presented with a set of data revealing the performance of Black and White high school students on achievement tests. Specifically, participants were told that 35% of Black students and 65% of White students passed the test. Separate summaries of performance of students at rich versus poor schools were also shown to participants and indicated that the type of school was correlated strongly with ethnicity and students' performance. Thus, *within* a specific type of school, there was no racial difference in the percentage of students who passed the achievement test.

During the experimental session, participants rated the intellectual ability of members of two groups (Group A, Group B) based on their performance on anagram tasks. The anagram task in this case was to rearrange 5 or 7 letters in order to produce a new word; the 7-letter anagram was much more difficult than the 5-letter anagram. Participants were then presented with four pieces of information: 1) the individual's group identity (A vs. B), 2) whether the individual solved or failed to solve the anagram, 3) the difficulty of the anagram (5-letter vs. 7-letter), and 4) the correct solution to the anagram. Participants were told that Group A members solved 10 of 25 anagrams and Group B solved 15 of 25 anagrams. However, Group A attempted more 7-letter anagrams and Group B attempted more 5-letter anagrams. Thus, on both 5-letter and 7-letter anagrams, Group A demonstrated a higher success rate than Group B. The results showed that participants who received the statistical training were more likely to consider the level of difficulty of the anagrams in making their ratings and thus more likely to accurately rate Group A as more capable of solving anagrams. Taken together, this study suggests that through the training of ANCOVA reasoning, individuals become better at considering the effect of covariance and confounding variables, and hence were less likely to form erroneous stereotypes toward certain groups.

Other strategies also show that stereotyping can be reduced through training. For example, one study examined the effect of associative learning in reducing the formation of illusory correlation. Associative learning refers to the learning of correlations between events (Murphy et al., 2011). The researchers argued that according to the associative learning perspective, the illusory correlation effect was the outcome of incomplete learning. That is, illusory correlation occurs because individuals are faster at learning more frequent than less frequent events and tend to overestimate correlations during the early phase of the learning curve. The researchers proposed that illusory correlation should diminish after sufficient training. In the training program, they adopted the same procedure as used by Hamilton and Gifford (1976), except that participants were presented with 10 blocks of 9 statements for a total of 90 statements and likeability judgments were recorded 10 times, once following each block. Results showed that during the initial blocks of training, likeability ratings of minority group were lower than those of majority group. However, final attitudes were more positive for both the majority and the minority group, such that there was no significant difference in judgments of the two groups. Based on these findings, Murphy et al. (2011) suggested that the emergence of illusory correlation depended on the amount of exposure that individuals had to the covariation/non-covariation of the target social groups with the frequent and infrequent behaviours in question. Thus, through increased learning and training, illusory correlation can be extinguished.

In summary, research indicates that stereotype formation can be inhibited at both behavioural and cognitive levels. Hence, developing an intervention that helps prevent stereotype formation at an early age is very important as it reduces children's vulnerability to the formation and endorsement of stereotypes.

Rational statistical learning. As discussed above, stereotype formation can be explained in part by erroneous statistical reasoning processes. Encouragingly, adults' stereotypical judgments can be eliminated through statistical training. What remains unclear, however, is whether children would similarly benefit from statistically-focused stereotype reduction training. Past research has shown that when making social judgments, adults are able to use statistical information, like sample size and variability (i.e., Kunda & Nisbett, 1986; Nisbett, Krantz, Jepson, & Kunda, 1983). This ability to use sample parameters in social judgment processes has earned human adults the label of "intuitive statisticians" (Peterson & Beach, 1967).

Of relevance to the present thesis, developmental psychologists have consistently shown that children are able to make statistical inferences at a very young age (i.e., Kirkham, Slemmer & Johnson 2002; Saffran, Aslin, & Newport, 1996; Xu & Tenenbaum, 2007). According to Xu and Kushnir (2013), infants learn about the world by applying general principles of Bayesian inference, in that they implicitly integrate prior knowledge with the incoming statistical information to estimate the posterior probabilities of the hypotheses. Moreover, Xu (2007) suggested that the Bayesian inference learning is a more promising candidate in understanding children's cognitive development than the associative learning, because it: 1) explicitly emphasizes the importance of prior knowledge, 2) explicitly emphasizes the importance of information input, and 3) allows individuals to combine these two to update their current belief or knowledge.

Supporting evidence for children's ability to make statistical inference showed that even infants are able to make inductive inferences about a randomly drawn sample using information about the population, and vice versa. For example, Xu and Garcia (2008) showed 6- and 8-month-old infants a box of ping-pong balls in which 80% of the balls were red and the other 20% were white. The infants saw an experimenter close her eyes and randomly draw out two samples of balls. One sample contained 4 red balls and one white ball. This sample was the probable sample because the proportion of red to white balls was consistent with the population. The other sample contained one red ball and 4 white balls. This sample was the improbable sample because the proportion of red to white balls was inconsistent with the population. Results showed that infants found the sample of 4 red balls and one white ball as more probable than the sample of one red ball and 4 white balls. These findings suggest that infants are able to make probability estimations in simple statistical-inference tasks and are sensitive to statistical relations between samples and populations.

In addition, researchers have found that infants are sensitive to sampling conditions when making generalizations from samples to population. For instance, Gweon, Tenenbaum, and Schulz (2010) demonstrated that 13- to 18-month-old infants considered the characteristics of a sample (i.e. random or nonrandom) when making generalizations from the sample to the population. In this study, children were presented with a transparent box that contained squeaky blue balls and inert yellow balls and saw an experimenter drawing a sample of balls from the box. Children were randomly assigned to 2 conditions. In one condition, the experimenter's sampling process did not provide any information that would help distinguish whether the experimenter was sampling from only the squeaky blue balls or sampling randomly from the whole box. In the other condition, the sampling process suggested that the experimenter was non-randomly selecting the squeaky blue balls. During the test phase, the experimenter took three blue balls from the box each at a time and squeezed the blue ball so that it squeaked. The experimenter then took an inert yellow ball from the box and gave it to the child. The variable of interest was the frequency of children squeezing the inert yellow ball. Results of this study showed that after seeing the experimenter's non-random sampling process, children assumed that the experimenter was purposely selecting squeaky blue balls and thus were less likely to squeeze the yellow ball.

Thus, Gweon et al.'s (2010) study suggested that after learning a sample was non-randomly selected from the population, children made fewer generalizations from the sample to population.

Lastly, developmental research shows that young children implicitly apply general principles of Bayesian inference by utilizing statistical evidence in the process of acquiring new concepts and revising old concepts. For example, Ma and Xu (2011) demonstrated that although 2-year-old children assumed that everyone would share the same preferences for different types of objects, they were able to interpret an experimenter's nonrandom sampling behaviour as an expression of preference that was inconsistent with their own. Thus, 2-year-old children revised their prior belief about the experimenter's preference and offered the toy that the experimenter non-randomly collected when asked to pass the toy that the experimenter liked better. Thus, young children are able to revise their prior beliefs after receiving statistical evidence that was inconsistent with their prior beliefs.

The Present Study

The goal of the present study was threefold. First, we investigated whether young children form stereotypic judgments about a minority group based on the perception of illusory correlation, as a replication of previous studies. Second, as mentioned above, Murphy et al. (2011) found that adults' illusory correlation could be significantly reduced after increased learning of the associations between the targets' behaviours and their group identity. Thus, we also examined whether children's perception of illusory correlation could be significantly reduced through the associative learning. Third, it has been found that statistical training can be used to significantly reduce the perception of illusory correlation among adults (Schaller et al., 1996), and that young children are able to learn from others and the surrounding environment by making rational statistical inference (e.g., Xu, 2007). In light of these findings, we examined whether statistical training could reduce children's perception of illusory correlation.

In addition to children's judgments about the minority group, we investigated, for the first time, the relationship between illusory correlation and children's playmate choice. Past research has consistently found that children prefer to play with peers of the same race and same gender (Finkelstein & Haskins, 1983; Fishbein & Imai, 1993). Moreover, Fishbein and Imai (1993) found that among 4- to 5.5-years-olds, there was a White-race playmate preference for boys and this preference was consistent with the evidence of positive-White racial biases in preschool-aged children (Aboud, 1988; Fishbein, 1992, Williams & Morland, 1976). Given the evidence that children's prior stereotypic beliefs about one social group could affect their playmate choice, we examined whether the perception of illusory correlation would lead children to avoid choosing a member of the minority group as a playmate.

Experiment 1

Experiment 1 examined whether children form stereotypic judgments about a minority group based on the perception of illusory correlation. Hamilton and Gifford's (1979) classic illusory correlation paradigm was adapted to fit the capacities of young children. Instead of reading descriptive sentences, each picture of the stimuli showed a child performing a desirable or undesirable behaviour. As in the classic illusory correlation paradigm, the depicted children were described as members of either a smaller (minority) or larger (majority) group. In this experiment, undesirable behaviours were less frequent than desirable behaviours. Thus, it was hypothesized that children would be more likely to perceive the numerically smaller minority group as less friendly due to the false association between the undesirable behaviours and the minority group.

Past research on illusory correlation in school-aged children included frequency estimation task (i.e., estimating the number of minority group members performing undesirable behaviours) and group attribution task (i.e., identifying the group to which the depicted child and behaviour belonged) (e.g., Johnston & Jacobs, 2003). However, given that the current study involved preschool-aged children, these two tasks were eliminated to suit the attention and memory span of the younger children. In this experiment, children were asked to rate the majority and the minority group's overall friendliness and to choose a playmate from either the majority or the minority group. It was hypothesized that children would give higher friendliness rating to the majority group and that they would be more likely to choose a playmate from the majority group.

Method

Participants. Participants included 48 children, 16 in each of three age groups: 5- to 6-year-olds (mean age = 5.5 years; 8 girls), 7- to 8-year-olds (mean age = 7.3 years; 8 girls), and 9- to 10-year-olds (mean age = 9.4 years; 13 girls). Among these participants, twenty-three were Caucasian, 14 Asian, 4 Mixed Race with Caucasian, 3 Other Race, 1 Arab, 1 Black, 1 Latino American, and one participant whose parent preferred not to disclose any ethnicity information. Six additional participants were excluded due to inattention (2), prior bias (2, children chose their playmate from the group wearing outfits in their favourite colour), and incompletion of the tasks (2). All participants were recruited from the Greater Toronto Area through a participant database and the Ontario Science Centre.

Materials. The main materials included two sets of 27 pictures: one set showing boys engaged in desirable and undesirable behaviours (for male participants), and the other set showing girls engaged in desirable and undesirable behaviours (for female participants).

Each picture showed an individual child performing one type of behaviour. The child in each picture was identified as a member of either the Green Group or the Yellow Group. The Green Group was depicted as wearing green outfits, and the Yellow Group was depicted as wearing yellow outfits. Among the 27 pictures, 18 depicted members of the majority group and 9 depicted members of the minority group (2:1). The terms “majority” and “minority” refer only to the relative sizes of the groups, so the group membership of each child was arbitrary. We did not use social categories such as gender and race in order to avoid potential confounding effects of existing stereotypic beliefs or the introduction of any potential stereotypes.

In the majority group, 12 members performed desirable behaviours and the other 6 performed undesirable behaviours. In the minority group, 6 members performed desirable behaviours and the other 3 performed undesirable behaviours. Thus, the ratio of the desirable to the undesirable behaviours across the two groups was constant at 2:1. The undesirable behaviours were infrequently occurring. The complete list of the desirable and undesirable behaviours is shown in Appendix A.

Figure 1. Examples of stimuli: Panel A (top) shows examples of desirable behaviours and Panel B (bottom) shows examples of undesirable behaviours.

Stimuli. The stimuli consisted of a series of PowerPoint slides on a laptop computer, including 5 slides for the introduction phase, 30 slides for the impression formation phase, 3 slides for the test phase, and 9 slides for the debriefing phase.

Introduction. During the introduction phase, an overview slide with all 27 pictures was first presented. The second slide demonstrated a 4-point Likert scale ranging from 1 (very mean) to 4 (very nice). Figure 2 shows this 4-point Likert scale. On this slide, four faces were used to represent the number of the scale. The first face was a sad face, meaning that the children were very mean. The second face was a slightly sad face, meaning that the children were kind of mean but not very mean. The third face was a slightly happy face, meaning that the children were kind of nice but not very nice. The fourth face was a happy face, meaning that the children were very nice. After the rating scale slide, two practice slides were shown. These two practice slides demonstrated two practice pictures depicting two individual children with red outfits. One picture depicted a child performing a desirable behaviour (helping teacher clean up toys), whereas the other depicted a child performing an undesirable behaviour (holding the umbrella for oneself and leaving another child in the rain). The 4-point Likert scale was shown at the bottom of each of the practice slides. Lastly, a slide with both practice pictures side by side was presented to allow participants to choose a playmate between the two children.

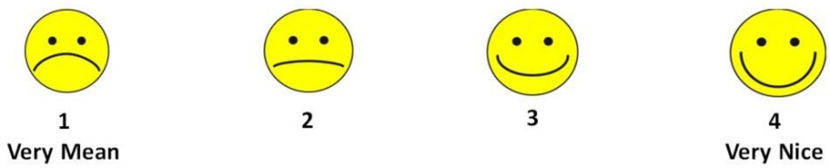

Figure 2. The 4-point Likert scale for friendliness rating.

Impression formation. During the impression formation phase, a slide of two prototype children was presented. The two prototype children were exactly the same except that one was wearing a green outfit, and the other one wearing a yellow outfit. Then, the 27 pictures (one per slide) as described above were shown to the child in one of two randomized orders, one at a time. Afterwards, the child was shown three review slides, with nine pictures on each slide, one picture at a time. The presentation order of the pictures on the review slides was the same as for the 27 individual pictures.

Figure 3. Examples of the prototype children.

Test. During the test phase, three slides were presented to participants. The first two slides were the rating slides, with the pictures of the prototype children from each group (one per slide). The 4-point Likert scale was shown on each slide under the picture of the prototype child. The last slide of the test phase demonstrated the two prototype children side by side in order to let participants choose a playmate between the two groups. The presentation order of each prototype child was counterbalanced.

Debriefing. During the debriefing phase, the 9 pictures that showed undesirable behaviours were presented to participants again each at a time (one per slide). These 9 pictures were presented in the same order as in the impression formation phase.

Three factors were counterbalanced across participants in each age group. First, the matching between the group identity (majority vs. minority) and the colour of the outfit (green vs. yellow) was counterbalanced. That is, the majority group was dressed in green and the minority group was in yellow for half of the participants, and vice versa for the other half of the participants. The order of introducing the two groups during the introduction phase was also counterbalanced; so was the order of rating the friendliness of each group during the test phase.

Procedure. Each child was tested individually either in the lab or in a quiet area at the Ontario Science Centre. The procedure consisted of four phases as detailed below.

Introduction phase. During the introduction phase, the experimenter told the child that he/she would view a series of pictures showing different children performing different behaviours, and that these children belonged to two groups, the Green Group and the Yellow Group. The child was also told that after viewing these pictures, he/she would be asked to rate the depicted children's friendliness by using "four faces" (i.e., the 4-point Likert scale as shown in Figure 2). The experimenter explained the scale to the child by pointing at each of the four faces and described its meaning. Afterwards, she asked the child to match each face with its corresponding meaning. Then, the child was asked to rate the friendliness of the two children shown in the practice slides. Lastly, the child was asked to choose a playmate between these two children.

Impression formation phase. During the impression formation phase, the experimenter presented the 30 slides as described earlier. When presenting the 27 pictures, the experimenter also provided verbal descriptions of the depicted child's group identity and behaviour in each picture. When presenting the three review slides, the child was asked to verbally report whether the depicted child in each picture was being mean or nice.

Test phase. During the test phase, the child was asked to rate the Green and the Yellow Group's overall friendliness on the same 4-point Likert scale as shown in Figure 2. In addition, the child was asked to choose a playmate from either the Green or the Yellow Group and to provide a reason for their playmate choice.

Debriefing. After the experiment, the researcher showed the child the pictures of the undesirable behaviours, described each picture, and told him/her that it was mean for the depicted children to engage in these behaviours.

Coding and reliability. The researcher coded children's responses during the experiment. For reliability, a trained research assistant who was blind to the hypothesis of the study coded a randomly selected 50% of the sample from video recordings. The two coders agreed 100% of the time.

Results

Preliminary analyses. To test the effects of presentation order, gender, and ethnicity (3: Caucasian, mixed with Caucasian, other race), three between-subjects Analysis of Variance

(ANOVA) were conducted. The results showed that there were no significant main effects on children's friendliness ratings (F ranges from 1.48 to 1.78, p ranges from .20 to .23) and playmate choice (χ^2 ranges from 0.49 to 3.56, p ranges from .39 to .83). Thus, these three variables were not included in the main analyses below.

Friendliness ratings. To investigate the effects of group identity and age on children's friendliness rating of the two groups, a 2 (group identity: majority, minority) x 3 (age group: 5-6, 7-8, 9-10) mixed-design ANOVA was conducted. The results showed that there was a significant main effect of group identity on children's friendliness rating, $F(1, 37) = 19.80$, $p < .001$, $\eta^2 = .31$, such that participants gave significantly higher ratings to the majority group than the minority group. The age effect and the interaction between age and group identity were not significant $F(2, 47) = 0.11$, $p = .90$, and $F(2, 42) = 0.48$, $p = .62$, respectively (see Table 1 for mean friendliness ratings by age group).

Table 1
Mean Friendliness Ratings by Age Group in Experiment 1 (*SD* in Parentheses)

Age	*N*	Majority	Minority	*t*-test
5- to 6-year-olds	16	3.25 (0.68)	2.38 (0.81)	**
7- to 8-year-olds	16	3.25 (0.78)	2.50 (0.73)	*
9- to 10-year-olds	16	3.13 (0.62)	2.63 (0.62)	**
Total	48	3.21 (0.68)	2.36 (0.72)	***

* $p < .05$, ** $p < .01$, *** $p < .001$

In addition, Fisher's exact test was conducted to examine the effect of age on whether children gave a higher or lower rating to the majority group, or gave equal ratings to both groups (see Table 2). Results showed that the age effect was not significant, $F(2,47) = 1.41$, $p = .91$. Therefore, the data were collapsed across age to examine whether participants gave a higher rating to the majority group at a level significantly above chance. Results of a Chi-square test showed that the number of the participants who gave a higher rating to the majority group

(32/48, 67%) was significantly greater than would be expected by chance (33.3%), χ^2 (2, 48) = 24, p < .001, Cramer's V = .71.

Table 2
Number of Children Who Gave a Higher Friendliness Rating to the Majority Group, Equal Ratings to Both Groups, and a Higher Rating to the Minority Group in Experiment 1

	Number of Children			
Age	Majority > Minority	Majority = Minority	Majority < Minority	Total
5- to 6-year-olds	11	2	3	16
7- to 8-year-olds	11	2	3	16
9- to 10-year-olds	10	4	2	16
Total	32	8	8	48

Playmate choice. To test the effect of age on children's playmate choice, Fisher's exact test was conducted given that more than 20% of the cells had expected counts less than 5. Table 3 demonstrates the number of children who chose their playmate either from the majority or the minority group across the three age groups.

Table 3
Playmate Choice by Age Group in Experiment 1

Age	Number of Children		
	Choosing Majority	Choosing Minority	Total
5- to 6-year-olds	13	3	16
7- to 8-year-olds	12	4	16
9- to 10-year-olds	11	5	16
Total	36	12	48

Results showed that the main effect of age was not significant, F (2, 47) = 0.73, p = .91. Therefore, the data were collapsed across age group to examine the effect of group identity on children's playmate choice. Results of the Fisher's exact test demonstrated that the number of children who chose a playmate from the majority group (36/48, 75%) was significantly greater than would be expected by chance (50%), χ^2 (1, 48) = 12, p < .001, Cramer's V = .5. Based on the odds ratio, children were three times more likely to choose their playmate from the majority group than from the minority group.

Reason for playmate choice. The reasons children provided for their playmate choices were coded as a categorical variable with three levels: 1) dispositional explanation (e.g., the chosen group was nice or nicer than the other group), 2) estimated frequency of the group members' behaviours (e.g., the chosen group has more members performing nice behaviours than the other group), and 3) other explanation (e.g., participants like the chosen group better). Table 4 demonstrates the number of children who provided dispositional, estimated frequency and other explanations for their playmate choice across all three age groups. A chi-square test showed that there was a significant age effect on children's reason for playmate choice, χ^2 (4, 48) = 11.39, p = .02, Cramer's V = .35 such that there were significantly more 5- to 6-year-olds provided other reasons for their playmate choice (z = 2.12, p = .04) than children from the other two age groups. However, the number of 5- to 6-year-olds who provided who provided the three types of reasons was not significantly different from chance level (33.33%), χ^2 (2, 16) = 0.88, p = .65. On the other hand, the number of 7- to 8-year olds who provided the three types of reasons was significantly different from chance level, χ^2 (2, 16) = 7.63, p = .02, Cramer's V = .49, such that 7- to 8-year-olds were more likely to provide dispositional explanations (10/16 = 62.5%, z = 2.02, p = .04) than chance level (33.33%). The number of 9- to 10-year-olds who provided the three types of reasons was also significantly different from chance level, χ^2 (2, 16) = 9.50, p = .01, Cramer's V = .54, such that 9- to 10-year-olds were more likely to provide reasons based on estimated frequency (10/16 = 62.5%, z = 2.02, p = .04) and less likely to provide other reasons (z = 2.31, p = .02) than chance level (33.33%).

Table 4
Number of Children Who Provided Dispositional, Estimated Frequency, and Other Reasons for Their Playmate Choice in Experiment 1

	Number of Children			
Age	Dispositional	Estimated Frequency	Other	Total
5- to 6-year-olds	7	4	5	16
7- to 8-year-olds	10*	5	1	16
9- to 10-year-olds	6	10*	0*	16
Total	23	19	6	48

* $p < .05$ as compared to chance

Discussion

Experiment 1 examined whether 5- to 10-year-old children formed stereotypical judgments about a minority group, on the basis of an illusory correlation that linked infrequent, undesirable behaviours with the minority group. The group membership (majority vs. minority) was determined by relative sizes of the groups and thus arbitrary. The results supported our hypothesis: Children gave a significantly higher rating of friendliness to the majority group than the minority group, and the number of children who gave a higher rating to the majority group was significantly greater than chance. Interestingly, there was no significant age effect on children's friendliness ratings. Consistent with previous findings (Johnson & Jacobs, 2003; (Primi & Agnoli, 2000), this lack of age difference suggests that the mechanism underlying the formation of illusory correlation is a process that might not undergo any developmental change, at least between the ages of 5 and 10 years.

In addition, the results showed that children were more likely to choose their playmate from the majority group than from the minority group. This finding provides the first indication that the perception of illusory correlation could further influence children's social interactions by affecting their playmate choice.

One might ask whether children's biased judgment of the minority group was due to the mere exposure effect rather than the perception of illusory correlation. That is, children viewed much more pictures depicting majority group members performing desirable behaviours and therefore perceived the majority group as friendlier. However, Hamilton and colleagues (1985) found that adults' biased judgment of the minority group was due to the overestimation of the degree to which members of the statistically infrequent (minority) group performed statistically infrequent (undesirable) behaviours rather than the degree to which members of the statistically frequent (majority) group performed statistically frequent (desirable) behaviours. In their experiment, the classic illusory correlation paradigm was used, except that participants were asked to complete a free recall task at the end of the procedure. The results showed that participants recalled a significantly higher proportion of the minority group/undesirable behaviour stimulus items than any of the other three types of sentences (Hamilton et al., 1985). Thus, these findings provided support for the interpretation that participants' biased judgment was formed upon biased recollection of the infrequent/distinctive information rather than the mere exposure effect.

To summarize, Experiment 1 added two new findings to the current literature. Previous studies suggest that children ages 6 and up form stereotypes on the basis of illusory correlation (e.g., Johnson & Jacobs, 2003; Primi & Agnoli, 2000). The present study provides the first evidence that 5-year-olds also perceive illusory correlation and make stereotypical judgments accordingly. In addition, this study is the first indication that the perception of illusory correlation can further affect children's playmate choice.

Experiments 2 and 3 were conducted to investigate two potential methods (i.e., associative versus rational statistical learning) that could be used to reduce the formation of illusory correlation among children, which might contribute to stereotype reduction.

Experiment 2

Experiment 2 aimed to examine whether the formation of illusory correlation could be reduced through increased associative learning. As mentioned earlier, Murphy and colleagues (2011) found that adults' illusory correlation could be eliminated through increased learning of the group-behaviour associations. In their study, adults were asked to learn 10 blocks of 9 group-behaviour associations. Following each block, they were asked to rate their likings of the majority versus the minority group. This learning process was called associative learning because the participants' judgments about the majority and the minority group were merely based on the learning of the group-behaviour associations (Murphy et al., 2011).

In Experiment 2, the same procedure from Murphy et al (2011) was used, but modified to suit young children's attention and memory span. Instead of learning 10 blocks of 9 descriptive sentences, 5- to 8-year-old children learned 3 blocks of 9 pictures of depicted characters performing desirable or undesirable behaviours. It was hypothesized that after two blocks of trainings on associative learning, there would be no significant difference in children's friendliness ratings of the majority versus the minority group. It was also hypothesized that children would choose their playmates randomly from the two groups because they would perceive the two groups as equally friendly.

Method

Participants. Participants included 35 children, sixteen 5- to 6-year-olds (mean age = 5.5 years, 8 girls) and nineteen 7- to 8-year-olds (mean age = 7.4 years, 13 girls). Among these participants, fourteen were Asian, 12 Caucasian, 3 Mixed Race with Caucasian, 2 Black, 1 Aboriginal, 1 other race, and one participant whose parent preferred not to disclose any ethnicity information. Five additional participants were excluded because of inattention (1), prior bias (3, children reported that they chose their playmate from the green group because green was their favourite colour), and experimenter error (1). All participants were recruited from the Greater Toronto Area through a participant database and the Ontario Science Centre.

Materials, stimuli, and procedure. The same materials and stimuli from Experiment 1 were used, except that the pictures were shown to participants in 3 blocks of 9 pictures. In each block, there were 6 members from the majority group (4 desirable and 2 undesirable behaviours) and 3 from the minority group (2 desirable behaviours and one undesirable behaviour).

The procedure was similar to that in Experiment 1, except that following each block of 9 pictures, the child saw a review slide displaying the 9 pictures just presented, one at a time, in the same order. When reviewing each picture, the child verbally reported whether the depicted child was mean or nice. After each review slide, the child was asked to rate the two groups' overall friendliness on the 4-point Likert scale. This process was repeated after each block.

Block 1 served as the baseline, and Blocks 2 and 3 constituted the associative-learning training blocks. Given that this experiment aimed to examine whether two blocks of training on associative learning could influence children's playmate choice and that children might change their choices after getting the same questions for multiple times, the child was asked to choose a playmate from the Green or the Yellow Group only at the end of the experiment.

Coding and reliability. Coding and reliability coding were done in the same manner as in Experiment 1. The two coders agreed 100% of the time.

Results

Preliminary analyses. Preliminary analyses indicated that across the 3 blocks, there was no significant main effect of presentation order, gender, or ethnicity on children's friendliness ratings (F ranges from 0.09 to 1.51, p ranges from .25 to .91) and playmate choices (χ^2 ranges from 0.29 to 2.02; *p* ranges from .86 to .15), so these three variables were not included in the main analyses.

Friendliness ratings. Table 5 shows children's mean friendliness ratings by age group across the three blocks. To investigate the effect of group identity and age on children's friendliness rating, three 2 (group identity: minority, majority) x 2 (age: 5-6, 7-8) mixed-design ANOVAs were conducted, one for each block. The results showed that there was a significant main effect of group identity on children's friendliness rating across all three blocks: Children gave significantly higher ratings to the majority group than to the minority group, $F(1, 34) = 14.53$, $p = .001$, $\eta^2 = .31$; $F(1, 34) = 17.64$, $p < .001$, $\eta^2 = .35$; and $F(1, 34) = 7.60$, $p = .01$, $\eta^2 = .19$, respectively. Across all three blocks, the difference in mean friendliness ratings of the majority and the minority group was significantly above chance (0), $t(34) = 3.83$, $p = .001$, Cohen's $d = .65$; $t(34) = 4.29$, $p < .001$, $d = .73$; and $t(34) = 2.67$, $p = .01$, $d = .45$, respectively. The main effect of age was not significant across all three blocks, $F(1, 34) = 0.67$, $p = .42$; $F(1, 34) = 1.04$, $p = .31$; and $F(1, 34) = 0.23$, $p = .63$, respectively; neither was the interaction between group identity and age, $F(1,31) = 0.20$, $p = .66$; $F(1,31) = 0.04$, $p = .85$; and $F(1, 31) = 1.10$, $p = .30$, respectively.

Table 5
Mean Friendliness Ratings by Age Group in Experiment 2 (*SD* in Parentheses)

	Age	*N*	Majority	Minority	*t*-test
Block 1	5- to 6-year-olds	16	3.50 (0.63)	2.50 (1.16)	*
	7- to 8-year-olds	19	3.26 (0.73)	2.47 (0.77)	*
	Total	35	3.37 (0.69)	2.49 (0.95)	**
Block 2	5- to 6-year-olds	16	3.06 (0.85)	2.44 (1.03)	*
	7- to 8-year-olds	19	3.32 (0.58)	2.63 (0.68)	***
	Total	35	3.20 (0.72)	2.36 (0.85)	***
Block 3	5- to 6-year-olds	16	3.50 (0.82)	2.56 (1.03)	*
	7- to 8-year-olds	19	3.16 (0.90)	2.74 (0.81)	n.s.
	Total	35	3.23 (0.87)	2.66 (0.91)	*

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 6 shows the number of children who gave a higher or lower friendliness rating to the majority group or equal ratings to both groups, by block and age group. Results of Fisher's exact test showed that across all three blocks, there were no significant age effect on whether children were more likely to give higher or lower ratings to the majority group, or give equal ratings to both group, F (2, 34) = 0.54, p = .80; F(2, 34) = 2.29, p = .32; F(2, 34) = 0.63, p = .90. Therefore, the data were collapsed across age to examine whether participants gave a higher rating to the majority group at a level significantly above chance. Results of a chi-square test showed that across all three blocks, the number of the participants who gave a higher rating to the majority group (21/35, 60%; 21/35, 60%; 22/35, 63%) was significantly greater than would be expected by chance (33.3%), χ^2 (1, 35) = 11.37, p < .001, Cramer's V=. 57; χ^2 (1, 35) = 12.74, p < .001, Cramer's V = .60; χ^2 (1, 35) = 13.77, p <.001, Cramer's V = .62.

Table 6
Number of Children Who Gave a Higher Friendliness Rating to the Majority Group, Equal Ratings to Both Groups, and a Higher Rating to the Minority Group Across 3 Blocks in Experiment 2

		Number of Children			
	Age	Majority > Minority	Majority = Minority	Majority < Minority	Total
Block 1	5- to 6-year-olds	10	4	2	16
	7- to 8-year-olds	11	4	4	19
	Total	21	8	6	35
Block 2	5- to 6-year-olds	10	3	3	16
	7- to 8-year-olds	11	7	1	19
	Total	21	10	4	35
Block 3	5- to 6-year-olds	11	2	3	16
	7- to 8-year-olds	11	4	4	19
	Total	22	6	7	35

Moreover, the number of participants who gave a higher friendliness rating to the majority group remained approximately the same across the three blocks, and this pattern was strikingly similar in both age groups.

Playmate choice. Chi-square test was applied to test the effect of age on children's playmate choice. Table 7 shows the number of children in each age group who chose their playmate from either the majority or the minority group. Results showed that there was no significant age effect, χ^2 (1, 35) = .18, p = .72. Thus, the data were collapsed across age to examine the effect of group identity on children's playmate choice. Among the 35 participants, 25 (71%) chose their playmate from the majority group, which was significantly above chance (50%), χ^2 (1, 35) = 6.43, p = .02, Cramer's V = .43.

Table 7
Playmate Choice by Age Group in Experiment 2

	Number of Children		
Age	Choosing Majority	Choosing Minority	Total
5- to 6-year-olds	12	4	16
7- to 8-year-olds	13	6	19
Total	25	10	35

Reason for playmate choice. Table 8 demonstrates the number of children who provided dispositional, estimated frequency and other reasons for their playmate choice across the two age groups. A chi-square test showed that there was no significant age effect on children's reasons for playmate choice, χ^2 (2, 35) = 3.55, p = .17. Therefore, the data were collapsed across age to examine whether participants provided each type of explanations for their playmate choice at a level significantly above chance. Results of a Chi-square test showed that the number of participants who provided the three types of reasons was significantly different from chance level, χ^2 (2, 35) = 6.40, p = .04, Cramer's V = .30, such that children were less likely to provide reasons other than dispositional attribution and estimated frequency (5/35 = 14.29%, z = - 1.95, p = .05) than chance level (33.33%).

Table 8
Number of children Who Provided Dispositional, Estimated Frequency, and Other Reasons for Their Playmate Choice in Experiment 2

	Number of Children			
Age	Dispositional	Estimated Frequency	Other	Total
5- to 6-year-olds	8	4	4	16
7- to 8-year-olds	9	9	1	19
Total	17	13	5*	35

* $p < .05$ as compared to chance

Discussion

Experiment 2 examined whether children's illusory correlation could be reduced through two blocks of associative learning. Results showed that after the training, participants still perceived the majority group as friendlier than the minority group. This pattern is consistent with Murphy et al.'s (2011) findings that after learning 27 trials of group-behaviour associations, adults still gave lower ratings to the minority group than the majority group. Although Murphy et al. (2011) demonstrated that there was no significant difference in the mean ratings at the end of the training (i.e., after learning 90 trials of group-behaviour associations), this training method presents limitations in reducing children's illusory correlation given that children had much smaller attention and memory span than adults do. Given that children were still more likely to choose their playmate from the majority group, the results of Experiment 2 suggested that two blocks of associative learning did not have a significant effect on children's judgment or their behaviours towards the two groups. Thus, Experiment 3 was conducted to investigate a different method that could be more efficient in reducing illusory correlation among children.

Experiment 3

The purpose of Experiment 3 was to examine whether the formation of illusory correlation could be reduced through rational statistical learning. As mentioned in the introduction, rational statistical learning refers to the process in which children learn about the world by combining prior knowledge with the input of statistical information in order to generate posterior beliefs (e.g., Xu & Kushnir, 2013). The design of Experiment 3 was similar to Experiment 2 with one major change. After viewing the pictures in Block 2, 5- to 8-year-olds were presented with a frequency summary table that demonstrated the breakdown of group identity (majority vs. minority) and the type of behaviour engagement (desirable vs. undesirable). This procedure was repeated after viewing the pictures in Block 3. Block 1 remained the same as in Experiment 2 and served as the baseline.

The presentation of the frequency summary table provided the opportunity for children to learn the accurate statistical information about the group-behaviour association, which is consistent with the rational statistical learning mechanism, such that the presentation of the frequency summary table provided statistical information and allowed children to integrate the incoming statistical evidence with their prior belief in order to generate posterior judgment. It was hypothesized that after learning the frequency summary table, children would perceive the majority and the minority group as equally friendly and choose their playmate between the majority and the minority group randomly.

Method

Participants. Participants included 32 children, sixteen 5- to 6-year-olds (mean age = 5.5 years, 9 girls) and sixteen 7- to 8-year-olds (mean age = 7.4 years, 8 girls). Among these participants, thirteen were Asian, 11 Caucasian, 3 Mixed Race with Caucasian, 2 Black, 1 Latino

American, 1 other race, and one participant whose parent preferred not to disclose any ethnicity information. Four additional participants were excluded because of inattention (2), prior bias (1, the child reported that he chose his playmate from the Green Group because green was his favourite colour), and difficulty understanding the instructions (1).

Materials, stimuli, and procedure. The pictures presented to participants in Experiment 3 are identical to those used in Experiments 1 and 2. The pictures were shown in 3 blocks of 9 pictures. The composition of pictures within each block was identical to the one in Experiment 2, such that the proportions of majority group to minority group members and desirable to undesirable behaviours were be preserved within each block.

Block 1 was conducted in the same manner as Experiment 2. In Blocks 2 and 3, the pictures were presented and reviewed in the same process as in Block 1. However, after reviewing the pictures, participants were presented with a frequency summary table as shown in Figure 4 that visually demonstrates the breakdown of the group identity (Green vs. Yellow) and the type of behaviour engagement (desirable vs. undesirable). Participants were asked to count the total number of members as well as the number of desirable and undesirable behaviours within each group. Participants were then asked to rate the Green and the Yellow Groups' overall friendliness on the 4-point Likert scale. At the end of the experiment, participants were asked to choose a playmate from either the Green group or the Yellow Group and to provide a reason for their playmate choice.

Coding and reliability. Coding and reliability coding were done in the same way as in the first two experiments. The two coders agreed 94% of the time. The 2 instances of disagreement were resolved by discussion.

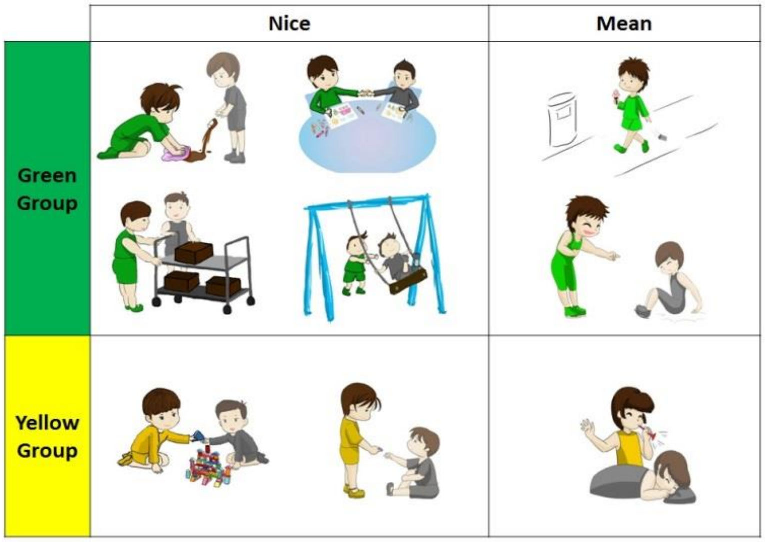

Figure 4. An example of the frequency summary table.

Results

Preliminary analyses. Similar to Experiment 2, mixed-design ANOVAs showed that across the 3 blocks, there was no significant gender, order, or ethnicity effect on participants' friendliness ratings (*F* ranges from .05 to 1.72, *p* ranges from .23 to .95) and playmate choices (χ^2 ranges from 0.56 to 15.42, *p* ranges from .23 to .76). These three variables were thus not included in the main analyses.

Friendliness ratings. In order to examine the effect of group identity and age on children's friendliness rating of the two groups, three 2 (group identity: majority, minority) x 2 (age group: 5-6, 7-8) mixed-design ANOVAs were conducted, one for each block. Please see Table 9 for children's mean friendliness ratings by age group across all three blocks. Results of the ANOVAs showed that there was a significant main effect of group identity on children's friendliness ratings in Block 1 only, F (1, 31) = 6.39, p = .02, η^2 = .17. In Block 2 and Block 3, there were no significant main effects of group identity, F(1, 31) = 3.04, p = .09; F(1, 31) = 1.34, p = .26. In Block 1, the difference in mean friendliness ratings of the majority and the minority group was significantly above chance (0), t (32) = 2.55, p = .02, d = .44. In Block 2 and Block 3, the difference in mean friendliness ratings was not significantly different from chance (0), t (32) = 1.77, p = .09, and t (32) = 1.18, p = .25, respectively. The main effect of age was significant in Block 1, F (1, 31) = 7.05, p = .01, η^2 = .19, but not in Block 2 or Block 3, F (1,31) = 0.00, p = .99; F (1, 31) = 2.03, p = .16. There was no significant interaction between group identity and age across all three blocks, F(1,28) = 0.03, p = .87; F(1,28) = 0.00, p = .96; and F(1, 28) = 0.13, p = .73, respectively.

Table 9
Mean Friendliness Ratings by Age Group in Experiment 3 (SD in Parentheses)

	Age	*N*	Majority	Minority	*t*-test
Block 1	5- to 6-year-olds	16	3.56 (0.63)	2.88 (0.96)	*
	7- to 8-year-olds	16	3.06 (0.68)	2.44 (0.63)	*
	Total	32	3.31 (0.69)	2.66 (0.83)	**
Block 2	5- to 6-year-olds	16	3.13 (0.62)	3.00 (0.86)	n.s.
	7- to 8-year-olds	16	3.19 (0.40)	2.88 (0.72)	n.s.
	Total	32	3.16 (0.52)	2.94 (0.80)	n.s.
Block 3	5- to 6-year-olds	16	3.19 (0.75)	2.94 (0.85)	n.s.
	7- to 8-year-olds	16	2.88 (0.81)	2.75 (0.58)	n.s.
	Total	32	3.03 (0.78)	2.84 (0.72)	n.s.

* $p < .05$, ** $p < .01$, *** $p < .001$

Table 10 demonstrates the number of children who gave a higher or lower friendliness rating to the majority group or equal ratings to both groups, by block and age group. Given that more than 20% of the cells have expected count less than 5, Fisher's exact test was applied to examine the effect of age on participants' ratings. Results showed that across all three blocks, there were no significant age effects on whether children were more likely to give higher or lower ratings to the majority group, or give equal ratings to both group, F (2, 31) = 0.84, p = .79; F (2, 31) = 1.37, p = .65; and F(2, 31) = 0.43, p = 1.00, respectively. Therefore, the data were collapsed across age to examine whether participants gave a higher rating to the majority group at a level significantly above chance.

Results of chi-square test showed that in Block 1, the number of participants who gave higher ratings to the majority group (19/32, 59%) was significantly greater than would be expected by chance (33%), χ^2 (1, 32) = 10.19, *p* = .001, V = .56. Whereas in Block 2, the number of participants who gave higher (11/32, 34%), lower (15/32, 47%) or equal (6/32, 18%) ratings to the majority and minority group were not significantly different from chance level (33%), χ^2 (1, 32) = 3.81, *p* = .10. Lastly, in Block 3, participants gave equal ratings to both groups (18/32, 56%) at a level significantly above chance, χ^2 (1, 32) = 8.31, *p* = .01, V = .51. This pattern is shown in Table 10 such that there was a decrease in the number of participants who gave higher ratings to the majority group; meanwhile there was an increase in the number of participants who gave equal friendliness ratings to the majority and the minority group. Such pattern is consistent with the changes in mean friendliness ratings across the 3 blocks.

Table 10
Number of Children Who Gave a Higher Friendliness Rating to the Majority Group, Equal Ratings to Both Groups, and a Higher Rating to the Minority Group in Experiment 3

		Number of Children			
	Age	Majority > Minority	Majority = Minority	Majority < Minority	Total
Block 1	5- to 6-year-olds	10	3	3	16
	7- to 8-year-olds	9	5	2	16
	Total	19	8	5	32
Block 2	5- to 6-year-olds	6	6	4	16
	7- to 8-year-olds	5	9	2	16
	Total	11	15	6	32
Block 3	5- to 6-year-olds	5	9	2	16
	7- to 8-year-olds	4	9	3	16
	Total	9	18	5	32

Playmate choice. Table 11 demonstrates the number of children who chose their playmate either from the majority or minority group across the 2 age groups. To examine the effect of age on children's playmate choice, 3 participants were excluded from the analysis because they refused to choose a playmate from either the majority or the minority group as they thought the two groups were equally friendly. In this analysis, a chi-square test was conducted given that none of the cells had expected counts less than 5. Results showed that there was no significant age effect on children's playmate choice, χ^2 (1, 32) = 0.85, p = .47. Thus, the data were collapsed across age to test for the effect of group identity on participants' playmate choice. Among the 29 participants, 14 (48%) chose their playmate from the majority group and 15 (52%) chose their playmate from the minority group. Results of the chi-square test demonstrated that this pattern was not significantly different from chance, χ^2 (1, 32) = 0.03, p = .86.

Table 11
Playmate Choice by Age Group in Experiment 3

	Number of Children		
Age	Choosing Majority	Choosing Minority	Total
5- to 6-year-olds	6	9	15
7- to 8-year-olds	8	6	14
Total	14	15	29

Reason for playmate choice. Table 12 demonstrates the number of children who provided dispositional, estimated frequency, and other reasons for their playmate choice across the two age groups. A chi-square test showed that there was no significant age effect on children's reasons for playmate choice, χ^2 (2, 32) = 1.14, p = .57. Therefore, the data were collapsed across age to examine whether participants provided reasons for their playmate choice based on the estimated frequency of the group members' behaviours at a level significantly above chance. Results of a chi-square test showed that the number of participants who provided dispositional reasons (15/32, 47%) was significantly greater than would be expected by chance (33.3%), χ^2 (2, 32) = 6.44, p = .04, Cramer's V = .32, such that children were less likely to provide reasons other than dispositional attribution and estimated frequency (4/32 = 12.50%, z = - 2.04, p = .04) than chance level (33.33%).

Table 12
Number of children Who Provided Dispositional, Estimated Frequency, and Other Reasons for Their Playmate Choice in Experiment 3

	Number of Children			
Age	Dispositional	Estimated Frequency	Other	Total
5- to 6-year-olds	7	8	1	16
7- to 8-year-olds	6	7	3	16
Total	13	15	4*	32

* $p < .05$ as compared to chance

Associative vs. Statistical Learning. In order to compare the effectiveness of associative and statistical learning in reducing illusory correlation, three 2 (group identity: majority vs. minority) by 2 (learning method: associative vs. statistical) mixed-design ANOVAs were conducted, one for each block. The results showed that for Block 1, there was no significant interaction between group identity and learning method on children's friendliness ratings, F (1, 65) = 0.58, p = .45. However, for Block 2, the interaction between group identity and learning method was statistically significant, F (1, 65) = 4.07, p = .04, η^2 = .06. Lastly, for Block 3, the interaction between group identity and learning method was non-significant, F (1, 65) = 2.48, p = .12.

Discussion

Experiment 3 examined whether children's stereotyping can be reduced through the training on rational statistical learning. More specifically, we investigated whether facilitating learning the accurate statistical information regarding the group identity and the type of behaviour engagement could inhibit the formation of illusory correlation among children. Results showed that after one block of statistical learning, children perceived the majority and the minority group as equally friendly. More interestingly, the results demonstrated a graded decrease in the number of children who gave higher ratings to the majority group across the three blocks. This pattern suggests that children were able to actively revise their prior judgment about the two groups as they were learning statistical information regarding the two groups. This finding is consistent with Ma and Xu (2011) such that 2-year-olds are able to revise their prior beliefs about an experimenter's toy preference through statistical learning.

Furthermore, there was no age difference in children's friendliness ratings after the 2 blocks of statistical training. This suggests that despite the developmental changes in children's mathematics ability (Cantlon & Li, 2014), 5- to 6-year-olds are as capable of learning statistical information as 7- to 8-year-olds are. The results are in line with recent findings of young children's ability of rational statistical learning such that even infants are able to make inductive inferences (Xu & Kushnir, 2013). Lastly, Experiment 3 demonstrated that after 2 blocks of training on statistical learning, children chose their playmate from the majority group at a level that was not significantly different from chance. The "intuitive" statistical learning ability could allow children not only to revise their prior belief but also to change their behaviour accordingly, so that in the end of the experiment, children chose their playmate between the majority and minority group randomly.

General Discussion

The three experiments reported here examined whether young children would make stereotypical judgments about the numerically smaller minority group based on the perception of illusory correlation, and whether the perception of illusory correlation could be reduced through training on associative learning and statistical learning. The results showed that 5- to 10-year-old children consistently perceived an illusory correlation between the numerically smaller minority group and the infrequently occurring, negative behaviour. However, the perception of an illusory correlation among 5- to 8-year-olds could be significantly reduced through the facilitation of statistical learning, but not associative learning.

More specifically, Experiment 1 demonstrated that there was no age effect on children's perception of illusory correlation, regardless of the known age differences in the verbal memory skills and other cognitive abilities of children between 5 and 10 years of age (Bjorklund & Coyle, 1995; Bjorklund & Muir, 1988). This lack of age difference suggests that the cognitive mechanism underlying the formation of illusory correlation might be unaffected by age, at least between the ages of 5 and 10 years. One possible explanation for this phenomenon is the fuzzy-trace theory proposed by Brainerd and Reyna (1992). According to this theory, individuals are able to extract verbatim and gist information when processing incoming stimuli. Verbatim and gist information are processed differently. Verbatim memory involves processing more exact traces of the incoming information. On the other hand, gist memory involves processing the patterns or relationships between the elements of the incoming information (fuzzy memory). Moreover, according to this theory, adults and children extract gist information in the same way and equally well such that both adults and young children can derive a variety of gist from the background inputs when processing information. For example, preschoolers can assess nominal gists such as “the As are most” and relational gists such as “more As than Bs” based on numerical inputs of mathematical problems that they will not solve correctly until some years later (Spinillo & Bryant, 1990). In light of this, illusory correlation might be explained as the consequence of people forming impressions of different social groups based on gist/fuzzy information rather than verbatim memory.

In addition, fuzzy-trace theory explains why adults often fail in simple tasks including class inclusion judgments (Agnoli, 1991) and simple probability judgments (Agnoli & Krantz, 1989; Tversky & Kahneman, 1983). This theory interprets adults’ poor performance on these tasks as a consequence of preference for making judgments based on fuzzy memories rather than accurate verbatim memories. Findings of Experiment 1 suggest that perhaps children have the same preference for fuzzy over verbatim memories as adults do; and hence, children consistently form illusory correlations regardless of the developmental change in their cognitive functions. Thus, future research might investigate the possibility of children’s preference for fuzzy over verbatim memories as accounting for the formation of stereotypic judgments about other people. If this hypothesis holds true, studies could then be implemented to test whether children’ illusory correlation can be inhibited when they are motivated to make impressions about other people based on the more accurate verbatim memory instead of gist information.

The results of Experiment 1 demonstrated the existence of illusory correlation among children by showing that children rated the minority group as less friendly than the majority group. However, the results did not directly demonstrate that young children overestimate the proportion of minority group members performing negative behaviours. Studies that explore the effect of illusory correlation in children who are 6 or older often include a frequency estimation task that assesses children’s impression of the relative proportions of positive and negative behaviours performed by each group (e.g., Johnston & Jacobs, 2003; Primi & Agnoli, 2002). However, this frequency estimation task was eliminated in this study because the current thesis investigated younger children (5-year-olds) who might not have the math ability to estimate the relative proportions and whose memory and attention span might be exhausted by merely viewing the 27 pictures. Although the current research did not include the frequency estimation task, children were asked to provide a reason for their playmate choice. More interestingly, almost forty percent of the children in Experiments 1 and 2 reported that they thought the majority group had more nice members (or fewer mean members) than the minority group, suggesting that children did, at least to some extent, rely on the proportions of positive and negative behaviours in each group when they were evaluating the two groups.

The results of Experiments 2 and 3 demonstrated that two blocks of training on rational statistical learning significantly reduced the formation of illusory correlation; however, such an inhibition effect was not obtained among children who received training on associative learning. These findings suggest that rational statistical learning might be a more effective method in reducing the formation of illusory correlation among children. Moreover, this pattern supports Xu's (2007) argument that Bayesian inference is a more promising candidate for understanding children's cognitive development than the associative learning theory because it explicitly emphasizes the importance of both prior knowledge and input information, and provides a way of combining the two. In Experiments 2 and 3, the mean friendliness rating difference between the majority and the minority group in Block 1 demonstrated children's prior knowledge about the two groups such that the majority group was perceived as friendlier than the minority group. However, it was only in Experiment 3 that children were given the opportunity to learn accurate statistical information about the two groups through learning the frequency summary table.

Consequently, it was only in Experiment 3 that the difference in mean friendliness ratings for the two groups became non-significant after two blocks of statistical learning. Therefore, by adopting the Bayesian inference perspective, the formation of illusory correlation can be explained as the consequence of participants' failure to combine the statistical input with their prior knowledge in order to generate posterior beliefs.

Although Murphy et al.'s (2011) training program showed that increased associative learning can significantly reduce illusory correlation among adults, it is perhaps a less feasible process for children as it requires a massive amount of trials in order to reduce the formation of illusory correlation (i.e., 90 group-behaviour associations were required for adult participants to give similar ratings to the majority and the minority group). In contrast, Johnston and Jacobs (2003) showed that illusory correlation is formed among second and fifth grade children after learning merely 18 group-behaviour associations. Thus, as suggested by Xu (2007), this type of exhaustive learning proposed by the associative learning theory might be contradictory to current knowledge about animal or human learning as animal or human learning is oftentimes more efficient than the associative learning demonstrated in Murphy et al.'s (2011) study.

Directions for Future Research

The current research demonstrates that children perceive illusory correlation between the minority group and infrequently occurring, negative behaviour, and that the formation of illusory correlation can be significantly reduced through training on rational statistical learning. The findings raise many other questions for future work. First, children's perception of illusory correlation revealed in this research is based on social stimuli depicting fictitious characters performing behaviours that are socially acceptable or unacceptable within their culture. Past research shows that the perception of illusory correlation can arise in non-social contexts as well.

For example, Primi and Agnoli (2002) used non-social stimuli including geometric figures with attributes of color and shape and found that 6- to 12-year-old children consistently perceived an illusory correlation between the infrequently occurring shape with the infrequently occurring colour. Moreover, a recent study by Lawson and Bower (2014) revealed the same phenomenon by using non-social stimuli involving two novel animal categories (okapis and coatimundis) and two novel biological properties ("plaxium blood" and "drotium blood"). Interestingly, the statistical training in Experiment 3 involves facilitating the process of organizing incoming information in a more statistically efficient manner, and such process is in line with the general principles of Bayesian/statistical learning. Given that Bayesian/statistical learning is an implicit mechanism that can be generalized to different modalities (e.g., language learning and visual perception; Kirkham et al., 2002), the training mechanism used in Experiment 3 should also be a basic and domain-general learning process. Therefore, future research could be conducted to examine whether the effect of facilitating statistical learning observed in the current research can be generalized to non-social contexts.

In addition, the present study assessed children's impression towards the two fictitious groups using only one friendliness scale. Although the results of Experiment 1 are very compelling such that children perceived the majority group as friendlier than the minority group, it would be interesting to include another scale that assesses children's perception of how unfriendly the two groups are. If children rate the minority group as more unfriendly than the majority group, such results would further complement the current findings of children's perception of illusory correlation between the minority group and the infrequently occurring negative behaviours. Moreover, past research that investigates illusory correlation in adults asked adults to rate the two groups in terms of 4 sets of traits. These 4 sets included positive and negative social traits (e.g., popular, sociable, vs. irritable, unhappy) and positive and negative intellectual traits (e.g., industrious, intelligent, vs. lazy, foolish). These studies found that when adults perceived the minority group to be associated with infrequently occurring undesirable behaviours, they gave higher ratings to the majority group for the positive social and intellectual traits and gave higher ratings to the minority group for the negative social and intellectual traits. Therefore, future research could investigate whether children's perception of illusory correlation could affect their impression of the minority group in terms of other social and intellectual traits.

Moreover, the formation of illusory correlation measured in Experiment 1 only assessed children's momentary impression and affective response towards two arbitrary groups. Thus, the degree to which illusory correlation leads to children's stable stereotypic beliefs in reality needs further examination. Similarly, although Experiment 3 showed that rational statistical learning could temporarily eliminate children's illusory correlation, it remains an open question whether this effect is long lasting. In addition to the long-term effect, another concern with reducing illusory correlation through statistical training is the degree to which the training can be generalized to real-world situations.

Although associative learning did not significantly reduce 5- to 8-year-old children's perception of illusory correlation, future research could be conducted to explore whether older children (e.g., 9- to 10- year-olds) might benefit from two blocks of associative learning. Similarly, given that 3-year-olds also perceive illusory correlation in non-social context (Lawson & Bower, 2014), future work can explore whether facilitating statistical learning can be used to reduce the formation of illusory correlation in non-social context among 3- to 4-year-olds.

Lastly, one possible explanation for the difference in effectiveness between associative and statistical learning is that through the presentation of the frequency summary table, children learnt a method to process incoming information in a statistically efficient manner, whereas associative learning relied more on children's implicit learning ability. Therefore, future research could investigate whether the effectiveness of associative learning in reducing illusory correlation would increase if children were explicitly told to pay attention to the number of desirable and undesirable behaviours performed by members of each group. In addition, given the possibility that children could acquire the ability to organize incoming information in a statistically efficient manner, future research could explore whether such ability could be generalized to the processing of various types of information.

Implications

Despite the open questions mentioned above, findings from this study may have important real-world implications. The results demonstrate that by providing accurate statistical information about two groups, children are able to revise their prior stereotypical beliefs and form more accurate impressions about these groups. Thus, the present findings might serve to inform interventions where parents and educators can help children reduce their stereotype formation, by explicitly emphasizing the importance of evaluating their prior knowledge about certain social groups on the basis of the incoming statistical evidence.

Past research showed that children's prejudice and stereotype could be successfully reduced through behavioural and cognitive-based interventions. For example, Aboud and Fenwick (1999) found that interventions facilitating in-class discussions about race and racial attitudes can reduce children's racial prejudice and that peers are valuable sources of this influence in the discussions. Moreover, in support of the intergroup contact theory, Pinquart, Wenzel, and Sörensen (2000) demonstrated that intergenerational attitudes could be significantly improved through joint activities between children and elderly adults, and even seven weeks after the end of the intervention, children's changes in attitudes toward elderly adults still remained significant. Lastly, Bigler and Liben (1992) showed that through cognitive training on multiple classification skills, children acquired the ability to process counter-stereotypic gender information, and thus, provided significantly more egalitarian responses on a subsequent measure of gender stereotyping. Given the confirming evidence that children's prejudice and stereotype could be reduced through interventions, the results of the present research therefore allow the possibility for new interventions that reduce children's stereotype.

According to Devine (1989), stereotype inhibition can be seen as a process of "breaking a bad habit" (*p*. 15). Therefore, stereotype reduction training should be provided at an early age to prevent the habitual use of stereotyping. Moreover, stereotype reduction training could lead to both short-term and long-term benefits for children's development, such that it has the potential to reduce the negative impacts of biased information processing and stereotype threat among children. More importantly, stereotype reduction training could facilitate children's skill acquisition in different counter-stereotypic domains (e.g., math and spatial memory), by promoting positive beliefs about and attitudes towards those domains. In this way, children's potential will be maximized such that their visions and choices will be broadened by breaking the boundaries set by those stereotypic beliefs.

References

Aboud, F. E. (1988). *Children and prejudice*. Oxford: Basil Blackwell.

Aboud, F. E., & Fenwick, V. (1999). Exploring and evaluating school‐based interventions to reduce prejudice. *Journal of Social Issues,* 55, 767-785. doi:10.1080/03601270050133883

Allport, G. W. (1979). *The Nature of Prejudice*. New York, NY: Basic Books.

Allport, G. W., & Kramer, B. M. (1946). Some roots of prejudice. *The Journal of Psychology, 22,* 9-39. doi: 10.1080/00223980.1946.9917293

Adomo, T. W., Frenkel-Brunswik, E., Levinson, D. J., & Sanford, R. N. (1950). *The authoritarian personality*. New York, NY: Harper.

Agnoli, F. (1991). Development of judgmental heuristics and logical reasoning: Training counteracts the representativeness heuristic. *Cognitive Development*, 6, 195–217.

Agnoli, F., & Krantz, D. H. (1989). Suppressing natural heuristics by formal instruction: The case of the conjunction fallacy. *Cognitive Psychology*, 21, 515–550.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice-Hall.

Bettelheim, B., & Janowitz, M. (1950). *Dynamics of prejudice*. New York, NY: Harper.

Biernat, M., & Vescio, T. K. (2002). She swings, she hits, she's great, she's benched: Implications of gender-based shifting standards for judgment and behaviour. *Personality and Social Psychology Bulletin*, *28,* 66-77. doi: 10.1177/0146167202281006

Bigler, R. S., & Liben, L. S. (1992). Cognitive mechanisms in children's gender stereotyping: Theoretical and educational implications of a cognitive-based intervention. *Child Development*, *63,* 1351–1363. doi: 10.1111/j.1467-8624.1992.tb01700.x

Bigler, R. S., & Liben, L. S. (1993). A cognitive–developmental approach to racial stereotyping and reconstructive memory in Euro-American children. *Child Development*, *64,* 1507–1518. doi: 10.1111/j.1467-8624.1993.tb02967.x

Bjorklund, D. F., & Coyle, T. R. (1995). Utilization deficiencies in the development of memory strategies. In F. E. Weinert & W. Schneider (Eds.), *Memory performance and competencies: Issues in growth and development* (pp. 161–180). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall: Remembering on their own. In R. Vasta (Ed.), *Annals of child development* (Vol. 5, pp. 79–123). Greenwich, CT: JAI.

Bodenhausen, G. V. (1990). Stereotypes as judgmental heuristics: Evidence of circadian variations in discrimination. *Psychological Science, 7*, 319-322. doi: 10.1111/j.1467-9280.1990.tb00226.x

Bodenhausen, G. V., & Lichtenstein, M. (1987). Social stereotypes and information processing strategies: The impact of task complexity. *Journal of Personality and Social Psychology*, *52,* 871-880. doi: 10.1037/0022-3514.52.5.871

Brainerd, C. J., & Reyna, V. F. (1992). Explaining "memory free" reasoning. *Psychological Science*, 3, 332–339.

Cantlon, J. F., & Li, R. (2013). Neural activity during natural viewing of Sesame Street statistically predicts test scores in early childhood. *PLoS biology*, 11, 1 – 13.

Campbell, D. T. (1967). Stereotypes and the perception of group differences. *American psychologist*, *22,* 817 - 829. . doi: 10.1037/h0025079

Carver, P. R., Yunger, J. L., & Perry, D. G. (2003). Gender identity and adjustment in middle childhood. *Sex Roles*, *49,* 95-109. doi: 10.1023/A:1024423012063

Chapman, L. J. (1967). Illusory correlation in observational report. *Journal of Verbal Learning and Verbal Behaviour*, *6,* 151-155. doi: 10.1016/S0022-5371(67)80066-5

Davidson, D., Cameron, P., & Jergovic, D. (1995). The effects of children’s stereotypes on their memory for elderly individuals. *Merrill–Palmer Quarterly, 41*, 70–90. http://www.jstor.org/stable/23087455

Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, *56*, 5 – 18. doi: 10.1037/0022-3514.56.1.5

Désert, M., Préaux, M., & Jund, R. (2009). So young and already victims of stereotype threat: Socio-economic status and performance of 6 to 9 years old children on Raven’s progressive matrices. *European journal of psychology of education*, *24*, 207-218. doi: 10.1007/BF03173012

Dovidio, J. F., ten Vergert, M., Stewart, T. L., Gaertner, S. L., Johnson, J. D., Esses, V. M., et al. (2004). Perspective and prejudice: Antecedents and mediating mechanisms. *Personality and Social Psychology Bulletin*, *30,* 1537-1549. doi: 10.1177/0146167204271177

Eagly, A. H., & Steffen, V. J. (1984). Gender stereotypes stem from the distribution of women and men into social roles. *Journal of personality and social psychology*, *46*, 735 – 754. doi: 10.1037/0022-3514.46.4.735

Eckes, T., & Trautner, H. M. (2000). The developmental social psychology of gender. Mahwah, NJ: Lawrence Erlbaum Associates.

Finkelstein, N. W., & Haskins, R. (1983). Kindergarten children prefer same color peers. *Child Development*, 54, 502 – 508.

Fishbein, H. D. (1992). The development of peer prejudice and discrimination in children. In J. Lynch, C. Modgil, & S. Modgil (Eds.), *Prejudice, polemic, or progress?* London: Falmer Press.

Fishbein, H. D., & Imai, S. (1993). Preschoolers select playmates on the basis of gender and race. *Applied Developmental Psychology*, 14, 303 – 316.

Fiske, S. T. (1989). Examining the role of intent: Toward understanding its role in stereotyping and prejudice. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 253-286). New York: Guilford Press.

Fiske, S. T, & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol.3, pp. 1-74). San Diego, CA: Academic Press.

Galinsky, A. D., & Moskowitz, G. B. (2000). Perspective-taking: Decreasing stereotype expression, stereotype accessibility, and ingroup favoritism. *Journal of Personality and Social Psychology, 78,* 708–724. doi: 10.1037/0022-3514.78.4.708

Gilbert, D. T, & Hixon, J. G. (1991). The trouble of thinking: Activation and application of stereotypic beliefs. *Journal of Personality and Social Psychology*, *60,* 509-517. doi: 10.1037/0022-3514.60.4.509

Goodman, M. E. (1952). *Race awareness in young children*. Cambridge, MA: Addison Wesley.

Gweon, H., Tenenbaum, J. B., & Schulz, L. E. (2010). Infants consider both the sample and the sampling process in inductive generalization. *Proceedings of the National Academy of Sciences*, *107,* 9066-9071. doi: 10.1073/pnas.1003095107

Hamilton, D. L., Dugan, P. M., & Trolier, T. K. (1985). The formation of stereotypic beliefs: Further evidence for distinctiveness-based illusory correlations. *Journal of personality and social psychology*, *48*, 5 – 17. doi: 10.1037/0022-3514.48.1.5

Hamilton, D. L., & Gifford, R. K. (1976). Illusory correlation in interpersonal perception: A cognitive basis of stereotypic judgments. *Journal of Experimental Social Psychology*, *12,* 392–407. doi: 10.1016/S0022-1031(76)80006-6

Hamilton, D. L., & Sherman, S. J. (1989). Illusory correlations: Implications for stereotype theory and research. In D. Bar-Tal, C. F. Graumann, A. W. Kruglanski, & W. Stroebe (Eds.), *Stereotyping and prejudice: Changing conceptions* (pp. 59–82). New York: Springer-Verlag.

Hastie, R., & Park, B. (1986). The relationship between memory and judgment depends on whether the judgment task is memory-based or on-line. *Psychological review*, *93*, 258-268. doi: 10.1037/0033-295X.93.3.258

Jacobs, J. E., & Narloch, R. H. (2001). Children's use of sample size and variability to make social inferences. *Journal of Applied Developmental Psychology*, *22*, 1–21. doi: 10.1016/S0193-3973(01)00086-7

Johnston, K. E., & Jacobs, J. E. (2003). Children's illusory correlations: The role of attentional bias in group impression formation. *Journal of cognition and* development, 4, 129–160 . doi:10.1207/S15327647JCD0402_01

Katz, P. A. (1976). The acquisition of racial attitudes in children. In P. A. Katz (Ed.), *Towards the elimination of racism* (pp. 125 – 154). New York: Pergamon Press.

Kim, H., & Baron, R. S. (1988). Exercise and illusory correlation: Does arousal heighten stereotypic processes? *Journal of Experimental Social Psychology*, *24*, 366-380. doi: 10.1016/0022-1031(88)90026-1

Kirkham, N., Slemmer, J., & Johnson, S. (2002). Visual statistical learning in infancy: Evidence for a domain general learning mechanism. *Cognition*, *83*, B35–B42. doi: 10.1016/S0010-0277(02)00004-5

Kit, K. A., Tuokko, H. A., & Mateer, C. A. (2008). A review of the stereotype threat literature and its application in a neurological population. *Neuropsychology Review*, *18,* 132-148. doi: 10.1007/s11065-008-9059-9

Kunda, Z., & Nisbett, R. E. (1986). Prediction and the partial understanding of the law of large numbers. *Journal of Experimental Social Psychology*, *22*, 339-354. doi: 10.1016/0022-1031(86)90019-3

Liben, L. S., & Bigler, R. S. (1987). Reformulating children's gender schemata. In L. S. Liben & M. L. Signorella (Eds.), *Children's gender schemata* (pp. 89-105). San Francisco: Jossey-Bass.

Liebert, R. M., Sobol, M. P., & Copemann, C. D. (1972). Effects of Vicarious Consequences and Race of Model upon Imitative Performance by Black Children. *Developmental Psychology*, *6*, 453 – 456. doi: 10.1037/h0020251

Ma, L., & Xu, F. (2011). Young children's use of statistical sampling evidence to infer the subjectivity of preferences. *Cognition*, *120*, 403-411. doi: 10.1016/j.cognition.2011.02.003

Macrae, C. N., Hewstone, M., & Griffiths, R. J. (1993). Processing load and memory for stereotype-based information. *European Journal of Social Psychology*, *23*, 77-87. doi: 10.1002/ejsp.2420230107

Macrae, C. N., Milne, A. B., & Bodenhausen, G. V. (1994). Stereotypes as energy-saving devices: A peek inside the cognitive toolbox. *Journal of Personality and Social Psychology*, *66*, 37-47. doi: 10.1037/0022-3514.66.1.37

McCabe, A. E., & Brannon, L. A. (2004). Effect of shifting standards on evaluations of white subtype members. *Journal of Social Psychology*, *144,* 91-92. doi:10.3200/SOCP.144.1.91-92

Miller, C. F., Lurye, L. E., Zosuls, K. M., & Ruble, D. N. (2009). Accessibility of gender stereotype domains: Developmental and gender differences in children. *Sex Roles*, *60,* 870-881. doi: 10.1007/s11199-009-9584-x

Mullen, B., & Johnson, C. (1990). Distinctiveness-based illusory correlations and stereotyping: A meta-analytic integration. *British Journal of Social Psychology*, *29*, 11–28. doi: 10.1111/j.2044-8309.1990.tb00883.x

Murphy, R. A., Schmeer, S., Vallee-Tourangeau, F., Mondragon, E., & Hilton, D. (2011). Making the illusory correlation effect appear and then disappear: The effects of increased learning. *The Quarterly Journal of Experimental Psychology*, *64*, 24-40. doi:10.1080/17470218.2010.493615

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, *43*, 747 - 759. doi: 10.1037/0012-1649.43.3.747

Nisbett, R. E., Krantz, D. H., Jepson, C., & Kunda, Z. (1983). The use of statistical heuristics in everyday inductive reasoning. *Psychological Review*, *90,* 339–363. doi: 10.1037/0033-295X.90.4.339

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2012). A threat in the classroom: Gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift für Psychologie/Journal of Psychology*, 220, 61 - 69. doi: 10.1027/2151-2604/a000097

Peterson, C. R., & Beach, L. R. (1967). Man as an intuitive statistician. *Psychological Bulletin*, *68*, 29 – 46. doi: 10.1037/h0024722

Piaget, J. (1965). *The child's conception of number*. New York, NY: Norton.

Pinquart, M., Wenzel, S., & Sörensen, M. (2000). Changes in attitudes among children and elderly adults in intergenerational group work. *Educational Gerontology*, 26, 523-540. doi: 10.1080/036012700050133883

Pratto, F, & Bargh, J. A. (1991). Stereotyping based upon apparently individuating information: Trait and global components of sex stereotypes under attention overload. *Journal of Experimental Social Psychology*, *27,* 26-47. doi: 10.1016/0022-1031(91)90009-U

Primi, C., & Agnoli, F. (2002). Children correlate infrequent behaviours with minority groups: A case of illusory correlation. *Cognitive Development*, *17,* 1105–1131. doi: 10.1016/S0885-2014(02)00076-X

Ramsey, J. L., & Langlois, J. H. (2002). Effects of the “beauty is good” stereotype on children's information processing. *Journal of Experimental Child Psychology, 8,* 320-340. doi: 10.1006/jecp.2002.2656

Reyna, V. F., & Ellis, C. J. (1994). Fuzzy-trace theory and framing effects in children's risky decision making. *Psychological Science*, 7, 275–279.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-monthold infants. *Science*, *274*, 1926–1928. doi: 10.1126/science.274.5294.1926

Schaller, M., Asp, C. H., Roseil, M. C., & Heim, S. J. (1996). Training in statistical reasoning inhibits the formation of erroneous group stereotypes. *Personality and Social Psychology Bulletin*, *22*, 829-844. doi: 10.1177/0146167296228006

Schaller, M., & O'Brien, M. (1992). " Intuitive Analysis of Covariance" and Group Stereotype Formation. *Personality and Social Psychology Bulletin*, *18*, 776-785. doi: 10.1177/0146167292186014

Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*, 336–356. doi: 10.1037/0033-295X.115.2.336

Sherman, S. J., & Corty, E. (1984). Cognitive heuristics. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 1, pp. 189-286). Hillsdale, NJ: Erlbaum.

Stangor, C, & Duan, C. (1991). Effects of multiple task demands upon memory for information about social groups. *Journal of Experimental Social Psychology*, 27, 357-378.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and intellectual test performance of African Americans. *Journal of personality and Social Psychology*, *69*, 797–811. doi:10.1037/0022-3514.69.5.797

Stewart, B. D., & Payne, B. K. (2008). Bringing automatic stereotyping under control: Implementation intentions as efficient means of thought control. *Personality and Social Psychology Bulletin*, *34*, 1332–1345. doi: 10.1177/0146167208321269

Stroessner, S. J., & Mackie, D. M. (1993). Affect and perceived group variability: Implications for stereotyping and prejudice. In D. M. Mackie & D. L. Hamilton (Eds.), Affect, cognition, and stereotyping: *Interactive processes in group perception* (pp. 63-86). San Diego, CA: Academic Press.

Taylor, S. E., & Crocker, J. (1981). Schematic bases of social information processing. In E. T. Higgins, C. P, Herman, & M. P. Zanna (Eds.), *Social cognition: The Ontario symposium* (Vol. 1, pp. 89-134). Hillsdale, NJ: Erlbaum.

Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental Psychology, 47,* 943-949. doi: 10.1037/a0024047

Tversky, A.,&Kahneman, D. (1983). Extensional versus intentional reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 91, 293–315.

Williams, J., & Morland, K. (1976). *Race, color, and the young child*. Chapel Hill: University of North Carolina Press.

Wilder, D. A., & Shapiro, P. (1989). Role of competition-induced anxiety in limiting the beneficial impact of positive behaviour by an outgroup member. *Journal of Personality and Social Psychology*, *56,* 60-69. doi: 10.1037/0022-3514.56.1.60

Xu, F. (2007). Rational statistical inference and cognitive development. *The innate mind: Foundations and the future*, *3,* 199-215.

Xu, F., & Garcia, V. (2008). Intuitive statistics by 8-month-old infants. *Proceedings of the National Academy of Sciences, 105,* 5012–5015. doi: 10.1073/pnas.0704450105

Xu, F. & Kushnir, T. (2013). Infants are rational constructivist learners. *Current Direction in Psychological Science*, *22*, 28-32. doi: 10.1177/0963721412469396

Xu, F., & Tenenbaum, J. (2007). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*, 288–297. doi: 10.1111/j.1467-7687.2007.00590.

Appendix A

List of Desirable Behaviours

1. Shares crayons with other children when drawing.
2. Helps another child build a tower block.
3. Shares lollipops with another child.
4. Helps another child get toy from a shelf that is too tall for the other child.
5. Helps another child clean up spills.
6. Helps parents shovel snow in the drive way.
7. Helps parents set the table for dinner.
8. Helps another child tie his/her shoelaces.
9. Read a picture book for a younger child.
10. Hands a band-aid to a child whose knee gets scraped.
11. Helps another child pick up a book.
12. Consoles a child who is crying.
13. Pushes the swing for a child.
14. Helps a child push a heavy cart.
15. Helps a child get up from a fall.
16. Picks up litter on a playground.
17. Holds the umbrella for another child when raining.
18. Helps a child put on a heavy backpack.

List of Undesirable Behaviours

1. Knocks over a child’s block tower.
2. Makes a lot of noise when a child is sleeping.
3. Destroys a child’s toy.
4. Opens another child’s birthday gift.
5. Blows out another child’s birthday candle.
6. Destroys another child’s drawing.
7. Throws garbage on the ground.
8. Jumps on the bed with dirty shoes on.
9. Laughs at a child who just falls over.