Gender Categorization in Infants and Children

Hong N. T. Bui

Senior Thesis in Psychology

Advisor: Karen Wynn

April 27, 2018

Abstract

Previous research indicates that, by as early as 5 months of age, infants categorize faces by gender. Additional research has examined that children as young as 4 years old spontaneously use gender as a cue to remember faces during a memory confusion paradigm. However, previous work has yet to establish whether these effects are driven by children's attention to low-level facial features that may be uncorrelated with gender. Using a habituation looking time paradigm, Study 1 examines 5-6-, 10-11-, and 15-19-month-old infants' categorization of faces that vary in gender typicality and control for overall perceptual similarity. Results reveal no significant difference in 5-6-, 10-11-, and 15-19 month-old infants' looking time to inconsistent vs. consistent gender test faces. Using a memory confusion paradigm, Study 2 examines whether 2-6 year old children spontaneously use gender to remember faces about unfamiliar people using the face stimuli from Study 1. Results reveal a marginally significant difference in automatic encoding of gender between older (5-6 years) and younger (2-4 years) children: during the recall task, older children were more likely to confuse people of the same gender with one another than to confuse males and females; younger children did not show this tendency. Lastly, both age groups reliably judged target faces low in gender typicality as more closely resembling of gender consistent faces than gender inconsistent faces. This was true despite the fact that the faces were equally perceptually distinct from the target face. These findings suggest that children but not infants use gender-specific features when categorizing faces.

Infants' and Children's Categorization of Gender

A great body of research indicates the significance of gender in social interactions. Gender is a salient category that is important evolutionarily and often drives social interactions (Cosmides, Tooby, & Kurzban, 2003; Dunham, Baron, & Banaji, 2016; Yee & Brown, 1994). Unfortunately, gender can easily become the basis of social stereotyping *because* it is a salient category that drives social interactions (Rutland, 1999). For example, by the age of 3, children already begin exhibiting same-gender preferences when choosing playmates (Maccoby, 2002). In comparison, children do not show social preferences based on race until around 5 years old (Kinzler & Spelke, 2011; Shutts, Banaji, & Spelke, 2010). It is important to note that these preferences occur in both implicit and explicit settings (Dunham, Baron, & Banaji, 2016). Additional support for the importance of gender in social categorization is from studies examining the impact of the social environment in shaping infants' gender preferences in their first months of life. By 3 months of age infants tend to visually orient more strongly to faces of one gender over the other, and this tendency is influenced by the gender of the infants' primary caregiver (Quinn, Yahr, & Kuhn, 2002). Specifically, infants raised primarily by a female caregiver express a preference for female faces, whereas infants raised primarily by a male caregiver express a preference for male faces. These findings suggest gender is an important category for infants and children, who show preferences for faces using gender cues. To examine the basis of social stereotyping and social preferences, it is crucial to examine initial contributing processes, such as categorization.

Previous literature is in strong support of gender categorization as a mechanism underlying gender stereotyping (Bigler & Liben, 2007; Martin & Halverson, 1981; Martin & Ruble, 2004). In order for children to display social preferences or stereotype awareness, regardless of the dimension they are based on, they must develop a reliable system of social categorization early on. In other words, they must have a reliable system that allows them to simplify their social world and generalize categories to new people (Bruner, 1957). Since gender is one of the first social categories children essentialize, or use to draw inferences about an individual's underlying traits, the onset of gender categorization is an important social category to examine in infancy. Furthermore, as children get older, they begin to receive more cultural input through language (e.g. "boys play baseball") about gender categories (Rhodes, 2013). Thus, measuring categorization of gender in preverbal infants, before the influence of language or culture, would inform understanding of the basic structural mechanisms used to evaluate others.

Research does show an ability to gender differentiate and categorize in infants (Fagan, 1976; Katz & Kofkin, 1997; Quinn, Yahr, & Kuhn, 2002; Younger & Fearing, 1999). Results from a habituation study by Younger & Fearing (1999) suggest that 10-month-olds, but not 7-month-olds, form categories based on gender. In this study, 10-month-olds generalized their habituation to faces that were gender-consistent with already habituated faces. That is, infants looked at new gender-consistent at the same rate as previously seen faces, indicating categorization. These infants also increased their attention to gender ambiguous faces, indicating that they did not generalize these new faces into the already established gender category. The authors argue that this trend was not seen in younger infants because at their age, they may have formed categories based on broad common features of faces, therefore, categorizing female and male faces into one category. However, a study by Fagan (1976) suggests the capacity to abstract characteristics from female and male faces is present by 7 months of age. After habituation to a male face, infants paid attention significantly more to the new female face compared to a new male face and compared to the habituated male face. Using the same habituation paradigm, other studies suggest that infants as young as 6 months of age categorize by gender (Katz & Kofkin, 1997; Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991).

In a study by Katz and Kofkin (1997), 6-month-old infants dishabituated to new faces within and outside the habituated gender category. However, infants looked significantly longer at new faces outside of the habituated gender category, compared to new faces within the established gender category, suggesting the use of gender cues to categorize. Young infants' ability to categorize by gender was not solely limited to visual stimuli. By 6 months of age, infants accurately matched faces with gender-consistent voices (Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991). However, this result was shown on only 50% of test trials. Evidence for more established intermodal gender knowledge was indicated at 12 months of age when infants matched gender typical faces with gender-consistent voices (Poulin-Dubois, Serbin, & Kenyon, 1994). In this study, infants were shown female and male faces from a modeling agency paired with recorded female and male voices. Infants looked longer at faces when they matched the voice of the same gender. Similar results were seen in 9-month-old infants, but only for highly stereotypical female faces and voices. The study reflects interesting gender differences in matching faces and voices by gender, but also suggests infant development of intermodal categorization by 9 months of age and stabilization by 12 months of age. This indicates that infants not only display preferences for faces based gender, but that they also hold specific expectations about gender consistency across various features.

Additional evidence for the role of gender categorization in infants comes from research examining object association to faces. Ten-month-old infants looked longer when an object previously paired with female is paired with a male, compared to when an object previously paired with a female is paired with another female (Levy & Haaf, 1994). These results suggested this early mechanism can act as the precursor to developing gender-based categories and stereotypes. Studies of this nature indicate that infants have early mechanisms that not only allow them to recognize gender differences, but also use these differences to represent faces categorically.

These results raise several developmental questions: Why do mechanisms for gender categorization develop so early in ontogeny? One explanation for the onset of gender categorization comes from an evolutionary standpoint. Evolutionarily, a mechanism that encodes gender would be advantageous for reproduction. Since societies consisted of people who varied in gender and age, mechanisms to categorize people by such characteristics would have been useful (Kinzler, Shutts, & Correll, 2010; Kurzban, Tooby, & Cosmides,2001). This corresponds with previous literature on the "big three" factors of social categorization: age, race, and gender (Brewer & Lui, 1989; Stangor, Lynch, Duan, & Glass, 1992; Macrae & Bodenhausen, 2000). When first meeting someone, these three factors, rather than say, clothing or whether or not someone wears glasses, are more likely to be remembered. That is, age, race, and gender are more likely to be automatically encoded (Stangor et al., 1992; Kruzban, Tooby & Cosmides, 2001). Such characteristics may be automatically encoded more strongly because they are highly informative, which is important in light of other characteristics that could be encoded. In addition, they provide information about an individual's underlying, stable characteristics (Stangor et al., 1992).

Automatic encoding of gender has been examined in children as young as 3 years old. Weisman and colleagues found automatic encoding of gender but not race in 4 and 5 year olds

(Weisman, Johnson, & Shutts, 2014). Using a well-developed "Who said What?" design (Taylor, Fiske, Etcoff & Ruderman, 1978), Weisman and colleagues (2014) measured participant errors in matching pictures of people with statements they previously said. Within-category errors were significantly greater than between-category errors, suggesting that children, as young as 4 years old, automatically encode gender. In connection to the evolutionary perspective of gender categorization, due to long-distances between societies, race would not have been as relevant of a characteristic, and this may be the reason why gender is automatically encoded more strongly than race (Cosmides, Tooby, & Kurzban, 2003; Weisman, Johnson, & Shutts, 2014). Thus, gender is one of the "big three" factors (Brewer & Lui, 1989; Stangor, Lynch, Duan, & Glass, 1992; Macrae & Bodenhausen, 2000) that adults and children automatically encode (Stangor et al., 1992; Cosmides, Tooby, & Kurzban, 2003; Kruzban, Tooby & Cosmides, 2001; Weisman, Johnson, & Shutts, 2014) and this may be in part due to its usefulness as an evolutionary cue (Stangor et al., 1992; Kruzban, Tooby & Cosmides, 2001).

Other researchers have proposed that children, in an attempt to make sense of the extensive information available to them about the world, categorize by features that are most salient, such as gender (Bigler, Jones, & Lobliner, 1997). This theory is coherent with findings that children focus on salient features during perceptual tasks (Livesley & Bromley, 1973) and Piaget's theory that children's thinking is heavily informed by the perceptual salience of people and objects (1970). However, as alluded to previously, social environment is also important in shaping children's evaluations and preferences (Quinn, Yahr, & Kuhn, 2002). Developmental intergroup theory (DIT) incorporates these two theories and posits that humans enter the world with a basic system that processes context-relevant information. As a result, gender may often be categorized because of its perceptual discriminability and the frequency of explicit gender labeling (Bigler & Liben, 2007; Liben, 2017).

While studies to date demonstrate that by the first year of life, infants are able to differentiate between the genders, develop preferences for one gender over another, make intermodal gender associations, and even attribute gender stereotypes through object association (Quinn et al., 2002; Levy & Haaf, 1994; Walker-Andrews, Bahrick, Raglioni, & Diaz, 1991), there are several key limitations which make interpreting their findings difficult. First, previous studies have not controlled for overall differences and similarities between face stimuli. For instance, one could argue that categorization abilities observed in previous studies are simply due to the visual similarity of stimuli within each category. For instance, two females may have characteristics that make them appear more similar to one another than either of them do to a male. Indeed, similarity of low-level features, which track with gender typical features, may drive infants' categorization from early on. In order to examine and measure whether infants are categorizing by gender per se, one needs to account for non-gendered similarities and differences between stimuli.

Previous attempts to control for similarity between faces have controlled for possible confounds such as hair length, clothing, and attractiveness (Fagan III & Singer, 1979; Leinbach & Fagot, 1993; Poulin-Dubois, Serbin, & Kenyon, 1994; Younger & Fearing, 1999). Poulin-Dubois, Serbin, & Kenyon (1994) and Leinbach & Fagot (1993) both used faces from magazines to control for attractiveness and found significant effects. Leinbach & Fagot (1993) demonstrated that when both hair and clothing were altered in stimuli to look more similar to one another, infants showed a significant decrease in recovery to a new face. That is, after habituation, they did not show as much interest in a face of the opposite gender, when both hair and clothing were altered. However, this was the case only when both clothing and hair were altered. Individual data suggest that 12-month old infants still categorized faces by gender, even when clothing or hair was altered such that the stimuli look more similar to one another. Additionally, Fagan III & Singer attempted to control for similarities across age and gender by presenting infants with similar stimuli such as a man and a baby, who differed in age but who are both bald and round-faced (1979). Their stimuli also included a white man and a white women who were siblings, and a black man and a black woman with the same hairstyle. Results suggested infants as young as 5 months could differentiate between male and female faces, even when the stimuli were similar in these ways. Although such results sound promising, as infants' categorization seems relatively stable and independent from incidental cues to gender, other research suggests the effects of other attributes such as baby-facedness or hue that could affect infants looking time (Kramer, Zebrowitz, Giovanni, & Sherak, 1995; Tarr, Kersten, Cheng, & Rossion, 2001). Thus, controlling for individual similarities and differences between stimuli is extremely important in light of previous findings.

Morphing techniques have allowed for careful control of stimuli. Face stimuli can be morphed to form gender ambiguous faces to examine categorization (Huart, Corneille, Becquart, 2005). Photographic images, rather than drawings of faces or computer generated images, provide rich textural information (Yamaguchi, 2000). With morphing techniques, this rich, realistic facial information can be kept while still controlling for stimuli factors. However, there is limited research on gender categorization that has used these precise morphing methods to ensure that similarities and differences are equated across facial stimuli (cf. Yamaguchi, 2000).

Yamaguchi (2000) used a morphing program to average faces and control for the distance in facial feature points between stimuli, as well as brightness. Stimuli were black and white photographs of 19-year-old Japanese faces (26 male, 26 female). These faces were used to make a hypermale and a hyperfemale face, as well as an average male, and average female face. Most importantly, the distance (in facial feature points) between the stimuli and the brightness were the same for all 4 stimuli. With these carefully controlled stimuli, Yamaguchi and colleagues found 8 month old infants, but not 6 month old infants, discriminated between female and male faces. Theses controls were extremely important to make sure any difference in looking time in infants was due to categorization by gender rather than categorization due to low-level differences between stimuli. However, this procedure still does not address the similarities and differences that may exist between the average male and the average female. The face stimuli used in this study were Caucasian. In addition, the face stimuli used were of higher quality, further contributing to the realism and ecological validity of the stimuli.

Face Stimuli

Stimuli Creation. Images from the MR2 Face database (Strohminger, Gray, Chituc, Heffner, Schein, & Heagins, 2016) were used to create stimuli for this study. This stimulus set was chosen because of its high-resolution, consistency of photography under standardized lighting conditions, and similarity in facial characteristics such as age, positioning, and mood. Controlling such factors is valuable if researchers want to measure the effects of the face itself, rather than the conditions in which the image was taken (Strohminger et al., 2016). As previously alluded to, maintaining naturalism in the facial stimuli is also extremely important. Thus, the MR2 database provided high resolution facial stimuli taken under very controlled conditions.

All of the 22 Caucasian male (N = 11) and female (N = 11) faces in the set were used. These images were edited using Psychomorph, a software designed to average faces based on specified points (Tiddeman, Stirrat, & Perrett, 2005; Sutherland, 2015). Using an averaging tool, the 22 faces were morphed to create a single androgynous face. This androgynous face was then rated by an independent group of 8 adults (4 male, M_{age}=20.5) who reported the extent to which it resembled each of the other 22 faces (see Figure 1 in Appendix). The gender typical female face in set A (F_{+a}) and gender typical male face in set A (M_{+a}) rated as least similar to the androgynous face were then used to create the final experiment faces. Two F_{+a} faces and 1 M_{+a} face was used to construct a lower-gender typical morph (F_{-a}) and 2 M_{+a} faces and 1 F_{+a} face was used to construct the other lower-gender typical morph (M_{-a}) (see Figure 2, set A in Appendix). Thus, individual similarities and differences were equal between these four faces. For the sake of brevity, F_{+a} and M_{+a} will be used to refer to the more gender-typical faces and F_{-a} and M_{-a} will be used to refer to the less gender-typical faces. This process was repeated using the next female and male face rated as least similar to the androgynous face to create a second set of faces (see Figure 2, set B in Appendix). These faces will be referred as M_{+b}, M_{-b}, F_{-b}, and F_{+b}.

Adult Ratings.

Masculinity/Femininity. Another independent group of 8 adults (4 male, M_{age} = 21.13) rated the final face sets for masculinity or femininity on a 7-point Likert scale (1 = extremely masculine…7 = extremely feminine; see Figure 3 in Appendix). Participants ratings of masculinity/femininity were reliable (α = .87). Results from an independent-sample t-test showed significant differences in the ratings of masculinity or femininity between the faces within each set in the predicted direction. That is, ratings for M_{+a} (M = 1.63, SD = 0.52) were higher in masculinity than ratings for M_{-a} (M = 2.38, SD = 0.74), $t(14) = -2.34$, $p < .05$, ratings for M_{-a} (M = 2.38, SD = 0.74) were higher in masculinity than ratings for F_{-a} (M = 3.75, SD = 0.89), $t(14) = -3.36$, $p < 0.01$, and ratings for F_{-a} (M = 3.75, SD = 0.89) were higher in masculinity than ratings F_{+a} (M = 5.38, SD = 0.74), $t(14) = -3.92$, $p < .01$. Differences in the ratings of masculinity or femininity were also significant for set B faces. Ratings for M_{+b} (*M* = 1.88, *SD* = 0.64) were higher in masculinity than ratings for M_{-b} (*M* = 2.63, *SD* = 0.52), $t(14) = -2.58$, $p < .05$. Ratings for M_{-b} (*M* = 2.63, *SD* = 0.52) were higher in masculinity than ratings for F_{-b} (*M* = 4.38, *SD* = 0.92), $t(14) = -4.70$, $p < 0.01$, and ratings for F_{-b} (*M* = 4.38, *SD* = 0.92) were higher in masculinity than ratings F_{+b} (*M* = 5.38, *SD* = 0.52), $t(14) = -2.69$, $p < .05$.

Attractiveness. Another independent group of 8 adults (4 male, $M_{age} = 22.25$) rated the new faces on attractiveness on a 7-point Likert scale (1 = extremely unattractive…7 = extremely attractive; see Figure 4 in Appendix). Participant ratings of attractiveness were reliable (α = .78). Results from an independent-sample t-test showed a significant difference in attractiveness of faces between set A (M = 3.03, SD = 0.97) and set B (M = 4.38, SD = 1.13), $t(62) = -5.12$, $p < .05$. Further independent-sample t-tests showed a significant difference in attractiveness of the masculine faces (M = 5.13, SD = 0.35) compared to both feminine faces (M = 4.00, SD = 0.92; M = 3.25, SD = 0.89), $t(7) = 3.81$; $t(7) = 6.36$, $p < 0.01$.

Study 1

Study 1 examines whether infants are capable of representing gender categories over the face set previously discussed that has been controlled for overall perceptual similarity. In this study, infants were habituated to Caucasian faces that were either male or female. At test, infants' looking time to two new individuals was measured, both of whom were equidistant from the original habituation face. However, only one of the faces crossed what adults would call the "gender boundary." Based on previous literature, it was hypothesized that infants as young as 5 months would be sensitive to the change and look longer at the face that crossed the gender category.

Method

Participants. A total of 24 infants completed at least one block of the procedure and were included in the data analysis. Fifty-eight percent of infants were White, 17% were Hispanic, 8% were Asian, and 17% were biracial. Eight were healthy full-term 5-6 month-old infants (3 males; *M* age = 0.50 years). Eight were 10-11 month-old infants (5 males; *M* age = 0.93 years). Nine were 15-19 month-old infants (7 males; *M* age = 1.46 years). In addition, 9 infants were not included in the analysis because they failed to complete at least one block due to fussiness or because of procedural error. Participants were recruited from a lab database, which includes infants from the New England region of the United States.

Procedure. Each infant was tested individually in a dimly lit room while seated on his or her caretaker's lap, facing a 42 inch video screen approximately 24 inches away. Parents wore taped glasses so they could not see the stimuli during testing as to reduce any bias to the infant. Parents were also told to refrain from speaking during the trials. Two experimenters tested each infant: a presenter who controlled the slide transitions and an observer who coded the infants' looking behavior. The observer sat behind a curtain about 24 inches away from the infant and observed their actions through an eyehole slit. The observer remained blind to the specific stimuli that were presented throughout the task.

Using a slide presentation created in Keynote, the experimenter presented the faces one at a time on the video screen. Looking time was recorded on a computer using jHab software (Casstevens, 2007). Looking time recording for each trial began after an initial 2 second sound used to capture the infant's attention. A trial ended once the infant looked away from the stimulus for 2 consecutive seconds or once the trial reached its 30 second maximum.

During the habituation phase, infants were repeatedly shown one of the low gender typicality faces, either male or female (e.g., M_{-a} or F_{-a}). The habituation phase continued until infants' average looking time over three consecutive trials decreased to 50% or less than their average looking time during the first three trials. Infants viewed a maximum of 14 habituation trials. A bell signaled the presenter to advance to the test trials once the infant had successfully habituated.

During test trials, infants were shown two new faces, each shown 3 times in alternation, for a total of 6 test trials. One test face match the gender of the habituation face and was high in gender typicality (e.g., M_{+a}); the other test face did not match the gender of the habituation face and was lower in gender typicality (e.g., F_{-a}). The gender (male vs. female) and degree of gender typicality (high vs. low) of the stimuli were counterbalanced across infants. Sounds and motions accompanied the stimuli. The task included a habituation phase followed immediately by a test phase. The habituation and test phase was repeated for a second block. In between the first and second blocks, parents were told they were halfway through the study and time was taken so the looking time coder could open up a new coding segment on jHab. Infants were randomly assigned to one of 16 sequences. Habituation face order, sounds, and motions were counterbalanced between sequences. This study design allowed infants to be tested on faces that were equidistant in similarities and differences from the habituated face, however, only one of the two test faces crossed what adults would typically call the male or female gender category.

Results

Averages of consistent (test trial face was gender-consistent with habituation face) and inconsistent (test trial face was gender-inconsistent with habituation face) looking time scores were calculated for 5-6 month olds, 10-11 month olds, and 15-19 month olds for all three test trials. For 5-6 month olds, results of an independent samples t-test showed a marginally significant difference in looking time between the inconsistent (*M* = 14.63 *,* *SD* = 8.76) and the consistent (*M* = 8.76*, SD* = 7.48) test trials, $t(22) = 1.75$, $p = 0.094$ for the first test trial set (see Figure 5). In the 2nd test trials, differences in looking time between the inconsistent (*M* = 5.46, *SD* = 4.52) and the consistent (*M* = 8.53, *SD* = 7.42) were not significant, $p > .05$. In the 3rd test trials, differences in looking time between the inconsistent (*M* = 7.31, *SD* = 8.33) and the consistent (*M* = 6.77, *SD* = 7.00) were also not significant, $p > .05$.

For 10-11- and 15-19-month olds, results of an independent samples t-tests showed no difference in looking times between the inconsistent and consistent test trials for each of the three trial sets (all *p*s >.05) (see Table 1).

		Inconsistent		Consistent		
Age	Trial	*Mean*	*SD*	*Mean*	*SD*	*p*
10-11 months	1	7.34	8.45	7.40	7.91	0.97
	2	7.95	7.13	5.60	5.25	0.39
	3	5.97	5.29	5.80	8.12	0.96
15-19 months	1	6.70	5.32	6.45	4.90	0.89
	2	5.81	8.12	5.06	5.50	0.77
	3	5.90	4.30	5.47	7.81	0.86

Table 1. Results of independent samples t-tests for 10-11 mos and 15-19 mos

Discussion

This study was designed in part to examine infant categorization of gender using stimuli carefully controlled for similarities and differences. It was hypothesized that infants as young as

5 months would categorize these carefully controlled faces by gender, in that they would show a significant difference in looking time when test faces were gender-inconsistent with habituation faces. The results were not consistent with the hypothesis that infants would look longer at the gender-inconsistent test faces. Infants aged 5-6 months, 10-11 months, and 15-19 months did not show significant differences in looking time in any of the 3 test trials. While acknowledging the small sample size and the difficulty of interpreting null results, these results may suggest that when similarities and differences between stimuli are controlled, infants fail to categorize faces by gender. Infants' mechanisms for detecting gender may need not be so finely tuned that they can detect such controlled differences between stimuli. In the world, people rarely have to detect similarities and differences with such specificity so perhaps such mechanism would not be necessary in infancy. This possibility is important to consider in light of previous research on infant gender categorization. More measures to control for similarities and differences in stimuli should be taken to ensure that results from such studies are due to the faces themselves, rather than low-level differences (Strohminger et al., 2016).

Another alternative explanation may be that infants did not perceive gender of faces as intended due to a limitation in the stimuli set. While adult ratings of the controlled stimuli showed ratings of gender typicality in the anticipated direction (high masculinity face < low masculinity face < low femininity face < high femininity face), mean ratings of masculinity/femininity tended to be towards the "masculine" side of the scale. For instance, the mean for Set A's low femininity face was 3.75, which is between a score of 3 ("a little bit masculine") and 4 ("neither masculine nor feminine"). This may be due in part to the fact that faces from this stimuli set were neutral and not smiling, which may have been interpreted by raters as a bit angry. Previous research indicates emotions such as anger may skew gender ratings towards masculinity (Hess, Adams, Grammer, & Kleck, 2009). To address this limitation, future studies should consider the effect of emotionality of faces on gender perception.

Study 2

Using the memory confusion paradigm (Taylor, Fiske, Etcoff & Ruderman, 1978), Study 2 examines whether children ages 2-6 years old are capable of automatically encoding gender over the face set previously discussed that has been controlled for overall perceptual similarity. In this study, children were shown four blocks of four faces each (either from set A or B) paired with the face's "favorite animal." Children were told to remember the pairings because later they would be asked to match the animals with the faces. We expected infants who are 11 month olds and older to be sensitive to the change and look longer at the face that crossed the gender category. Based on previous literature, it was hypothesized that children as young as 3 years old would automatically encode by gender (Weisman, Johnson, & Shutts, 2015). That is, children would make more within-gender (e.g. confusing F_{+a} and F_{-a} or M_{+a} and M_{-a}) errors than between-gender (e.g. confusing F_{+a} and M_{-a}) when matching animals with faces.

Method

Participants. Twenty healthy full-term 2-4 year old children (5 males, 15 females; *M* age = 3.74 years) and 10 healthy full-term 5-6 year old children (3 males, 7 females; *M* age = 5.78 years) completed the procedure. In addition, 12 children were not included in the analysis due to piloting (n = 8), or fussiness (n = 4). Children were recruited through the same database and recruitment methods in Experiment 1.

Materials. Thirty-two trial sequences were available to use for the task. Order of familiarization faces shown and order of matching faces was counterbalanced. Tasks featured the same face stimuli from Study 1, shown on a laminated card with a velcro spot where children can match the animals to the faces.

Procedure. Participants were tested in the lab. Parents who accompanied their children into the testing room were asked to wear taped sunglasses and to refrain from speaking so as not to bias their children's answers. The experimenter informed the participants that they would see a series of people and their favorite animals. They were to pay really close attention because the experimenter would later ask the participant to match the animals with the faces. The task consisted of four blocks, each block consisting of a familiarization segment followed by a test segment. During the familiarization phase, participants saw four face stimuli and their favorite animals. The order the animals were shown in remained the same while the order of faces shown was counterbalanced. The face and animal were introduced as “Do you see this person? This person’s favorite animal is a cow.” Participants saw each pairing for a total of 12 s. During the test phase, participants were shown the four faces (shown in a row, order counterbalanced) and the four animals (randomly clustered in a cross shape) and were asked to match the animals to the faces. The same four faces were shown each time during the blocks. Lastly, children were shown one of the low typical faces (e.g. M_{-a}) and the two faces equidistant from that face (e.g. M_{+a} and F_{-a}) and were asked the explicit question, “Do you see this face (M_{-a}), does this face look more like this face (M_{+a}) or this face (F_{-a})?” Children’s answers during the test phase and the explicit question were recorded on a coding sheet.

Scoring. Adapting previous automatic encoding research (Taylor et al., 1978; Weisman, Johnson, & Shutts, 2015), an adjusted error difference score was calculated, using participants (total within-category errors) - (total between-category errors/2). Errors due to confusion of a female’s favorite animal to another female’s favorite animal (or of two males’ favorite animals) were within-category errors, whereas errors due to confusion of a female and a male's favorite animals were between-category errors. Between-category errors are twice as likely to occur compared to within-category errors, hence the dividing by 2. Scores could range from -8 (between-category errors in all four blocks) to 8 (within-category errors in all four blocks).

Results

Preliminary analysis indicated average auto-encoding score did not differ significantly for children ages 2 years, 3 years, and 4 years, thus, data was combined for children ages 2-4 years and 5-6 years. Results from an independent samples t-test showed a marginally significant difference in auto-encoding scores for 2-4 year olds ($M = -.55$, $SD = 3.33$) and 5-6 year olds ($M = 2.00$, $SD = 3.71$), $t(28) = -1.90$, $p = 0.07$ (see Figure 6). Results from an independent samples t-test showed a significant difference in number of correct matches between 2-4 year olds ($M = 3.80$, $SD = 2.35$) and 5-6 year olds ($M = 5.90$, $SD = 2.56$), $t(28) = -2.24$, $p < .05$. For 2-4 year olds, results from a one-sample t-test showed no difference in the number correct matches ($M = 3.80$, $SD = 2.35$) compared to chance, $t(28) = -0.38$, $p > .05$. For 5-6 year olds, results from a one-sample t-test showed a significant difference in the number of correct matches ($M = 5.90$, $SD = 2.56$) compared to chance, $t(28) = 2.35$, $p < .05$. A binomial test indicated that the proportion of 2-4 year old children (.75) making an explicit within-gender judgement was higher than the expected .50, $p < .05$. This was also true for the proportion of 5-6 year old children (.90) making the explicit within-gender judgement, $p < .05$ (see Figure 7).

Discussion

Study 2 was designed to examine children's automatic encoding of gender using stimuli carefully controlled for similarities and differences. It was hypothesized that children as young as 4 years would automatically encode gender, indicating that they use it as a social cue when learning facts about new people. There was a marginally significant difference in auto-encoding scores between 5-6-year-olds and 2-4-year-olds. That is, 5-6-year-olds made significantly more within-gender errors in the matching task. Compared to previous research examining children's automatic encoding of gender, the earliest sign of automatic encoding was in 5 year olds, whereas previously it was in 4 year olds (Weisman et al., 2015). This could suggest that automatic encoding in 4 year olds was due to recognition of low level differences between the non-controlled stimuli.

In addition, 5-6-year-olds performed significantly better than chance at correctly matching the animals with the respective faces. Taking this finding with the marginally significant finding of gender automatic encoding, 5-6-year-olds performed better at the task but also automatically encoded gender more than 2-4-year-olds. This may indicate that 5-6-year-olds use gender as a cue to learning new facts about people, but 2-4-year-olds do not. However, one possibility is that 2-4-year-olds simply did not understand the task well enough and matched faces randomly. Upon examining the order in which animals were placed down with faces, some children simply remembered the order of the animals and matched them with the faces, even though the order of the test faces was not the same as the order of familiarization with the animals. Thus, 2-4-year-olds may not have fully understood the task.

Another possibility is that 5-6-year-olds were simply more motivated to complete the task correctly, as suggested in the higher number of correct matches. Two-four year olds may have been less motivated to correctly complete the task, thus, performing at chance level. In contrast to this possibility, a few 2-4-year-old participants did ask "Is this correct?" a number of times during the study. In future studies, it will be important to keep in mind children's motivations to complete the task.

However, these results are interesting in light of the fact that the proportion for both 2-4 year olds and 5-6 year olds who made the explicit within-gender choice was significant. That is, when asked which face looked more similar to either the morphed low femininity or morphed low masculinity face, both groups of children chose the within-gender category face. However, 2-4 year olds did not necessarily use this information when matching. One possibility is that the explicit question asking which face looks more *similar* draws attention to specific similarities and thus more to the gender of the face so children were more likely to make the within-gender choice. In contrast, when simply matching the animals with faces and no reference was made to any characteristic of the faces, 2-4 year olds simply did not focus on similarities and differences between faces.

General Discussion

Taken together, the findings from the current experiments suggest that children but not infants may categorize by gender when stimuli are controlled for similarities and differences in facial stimuli. This may be due to the fact that although gender is an important category to detect in infancy, infants need not detect gender with such specificity. Additionally, 5-6-year-old children indicate the use of gender as a cue to learn facts about unfamiliar faces. Although this is not seen in 2-4-year-olds, when asked an explicit question about similarities between faces, they (along with 5-6-year olds) reliably judged faces low in gender typicality as resembling the within-gender face rather than the cross-gender face, even though the similarities between these faces was equal. These results are important to consider in light of previous research on infant and child categorization of gender. Measures to control for similarities and differences in facial stimuli are crucial to distinguish gender categorization from differentiation of low-level features uncorrelated with gender.

References

Bigler, R. S., & Liben, L. S. (2007). Developmental intergroup theory: Explaining and reducing children's social stereotyping and prejudice. *Current Directions in Psychological Science*, *16*(3), 162-166.

Bigler, R. S., Jones, L. C., & Lobliner, D. B. (1997). Social categorization and the formation of intergroup attitudes in children. *Child development*, *68*(3), 530-543.

Brewer, M. B., & Lui, L. N. (1989). The primacy of age and sex in the structure of person categories. *Social cognition*, *7*(3), 262-274.

Bruner, J. S. (1957). On perceptual readiness. *Psychological review*, *64*(2), 123.

Casstevens, R. M. (2007). jHab: Java habituation software (version 1.0. 2)[computer software]. *Chevy Chase, MD*.

Cosmides, L., Tooby, J., & Kurzban, R. (2003). Perceptions of race. *Trends in cognitive sciences*, *7*(4), 173-179.

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science*, *19*(5), 781-789.

Fagan III, J. F. (1976). Infants' recognition of invariant features of faces. *Child Development*, 627-638.

Fagan III, J. F., & Singer, L. T. (1979). The role of simple feature differences in infants' recognition of faces. *Infant Behavior and Development*, *2*, 39-45.

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men?. *Journal of Vision*, *9*(12), 19-19.

Huart, J., Corneille, O., & Becquart, E. (2005). Face-based categorization, context-based categorization, and distortions in the recollection of gender ambiguous faces. *Journal of Experimental Social Psychology*, *41*(6), 598-608.

Katz, P. A., & Kofkin, J. A. (1997). Race, gender, and young children. *Developmental psychopathology: Perspectives on adjustment, risk, and disorder*, *21*, 51-74.

Kinzler, K. D., & Spelke, E. S. (2011). Do infants show social preferences for people differing in race?. *Cognition*, *119*(1), 1-9.

Kinzler, K. D., Shutts, K., & Correll, J. (2010). Priorities in social categories. *European Journal of Social Psychology*, *40*(4), 581-592.

Kramer, S., Zebrowitz, L. A., San Giovanni, J. P., & Sherak, B. (1995). Infant preferences for attractiveness and babyfaceness. *Studies in perception and action III*, 389-392.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences*, *98*(26), 15387-15392.

Leinbach, M. D., & Fagot, B. I. (1993). Categorical habituation to male and female faces: Gender schematic processing in infancy. *Infant Behavior and Development*, *16*(3), 317-332.

Levy, G. D., & Haaf, R. A. (1994). Detection of gender-related categories by 10-month-old infants. *Infant Behavior and Development*, *17*(4), 457-459.

Liben, L. S., & Bigler, R. S. (2017). Understanding and undermining the development of gender dichotomies: The legacy of Sandra Lipsitz Bem. *Sex Roles*, *76*(9-10), 544-555.

Livesley, W. J., & Bromley, D. B. (1973). Person perception in childhood and adolescence.

Maccoby, E. E. (2002). Gender and group process: A developmental perspective. *Current directions in psychological science*, *11*(2), 54-58.

Macrae, C. N., & Bodenhausen, G. V. (2000). Social cognition: Thinking categorically about others. *Annual review of psychology*, *51*(1), 93-120.

Martin, C. L., & Halverson Jr, C. F. (1981). A schematic processing model of sex typing and stereotyping in children. *Child development*, 1119-1134.

Martin, C. L., & Ruble, D. (2004). Children's search for gender cues: Cognitive perspectives on gender development. *Current directions in psychological science*, *13*(2), 67-70.

Oakes, L. M. (2010). Using habituation of looking time to assess mental processes in infancy. *Journal of Cognition and Development*, *11*(3), 255-268.

Piaget, J. (1970). Science of education and the psychology of the child. Trans. D. Coltman.

Poulin-Dubois, D., Serbin, L. A., Kenyon, B., & Derbyshire, A. (1994). Infants' intermodal knowledge about gender. *Developmental Psychology*, *30*(3), 436.

Quinn, P. C., Yahr, J., Kuhn, A., Slater, A. M., & Pascalis, O. (2002). Representation of the gender of human faces by infants: A preference for female. *Perception*, *31*(9), 1109-1121.

Rhodes, M. (2013). How two intuitive theories shape the development of social categorization. *Child Development Perspectives*, *7*(1), 12-16.

Rutland, A. (1999). The development of national prejudice, in- group favouritism and self-stereotypes in British children. *British Journal of Social Psychology*, *38*(1), 55-70.

Shutts, K., Banaji, M. R., & Spelke, E. S. (2010). Social categories guide young children's preferences for novel objects. *Developmental Science*, *13*(4), 599-610.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. *Journal of Personality and Social Psychology*, *62*(2), 207.

Strohminger, N., Gray, K., Chituc, V., Heffner, J., Schein, C., & Heagins, T. B. (2016). The MR2: A multi-racial, mega-resolution database of facial stimuli. *Behavior research methods*, *48*(3), 1197-1204.

Sutherland, C. A. M. (2015). A basic guide to Psychomorph. *University of York. Available online at: https://dlib. york. ac. uk/yodl/app/home/detail*.

Tarr, M. J., Kersten, D., Cheng, Y., & Rossion, B. (2001). It's Pat! Sexing faces using only red and green. *Journal of Vision*, *1*(3), 337-337.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of personality and social psychology*, *36*(7), 778.

Tiddeman, B. P., Stirrat, M. R., & Perrett, D. I. (2005, September). Towards realism in facial image transformation: Results of a wavelet mrf method. In *Computer Graphics Forum* (Vol. 24, No. 3, pp. 449-456). Blackwell Publishing, Inc.

Walker-Andrews, A. S., Bahrick, L. E., Raglioni, S. S., & Diaz, I. (1991). Infants' bimodal perception of gender. *Ecological Psychology*, *3*(2), 55-75.

Weisman, K., Johnson, M. V., & Shutts, K. (2015). Young children's automatic encoding of social categories. *Developmental science*, *18*(6), 1036-1043.

Yamaguchi, M. K. (2000). Discriminating the sex of faces by 6-and 8-mo.-old infants. *Perceptual and motor skills*, *91*(2), 653-664.

Yee, M., & Brown, R. (1994). The development of gender differentiation in young children. *British Journal of Social Psychology*, *33*(2), 183-196.

Younger, B. A., & Fearing, D. D. (1999). Parsing items into separate categories: Developmental change in infant categorization. *Child Development*, *70*(2), 291-303.

Appendix

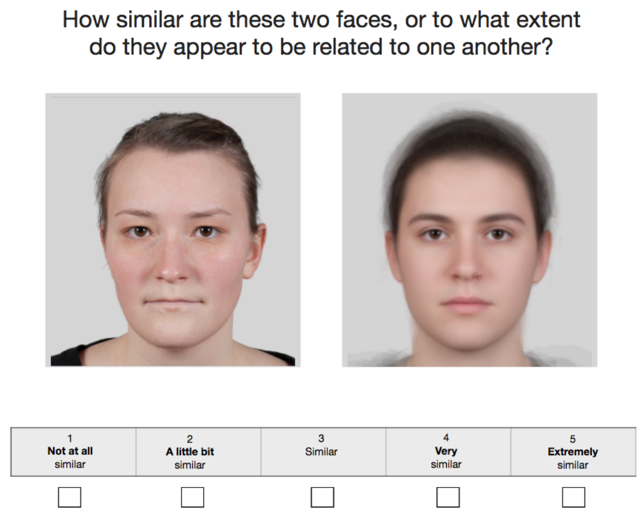

Figure 1. Rating questionnaire of individual compared to androgynous face

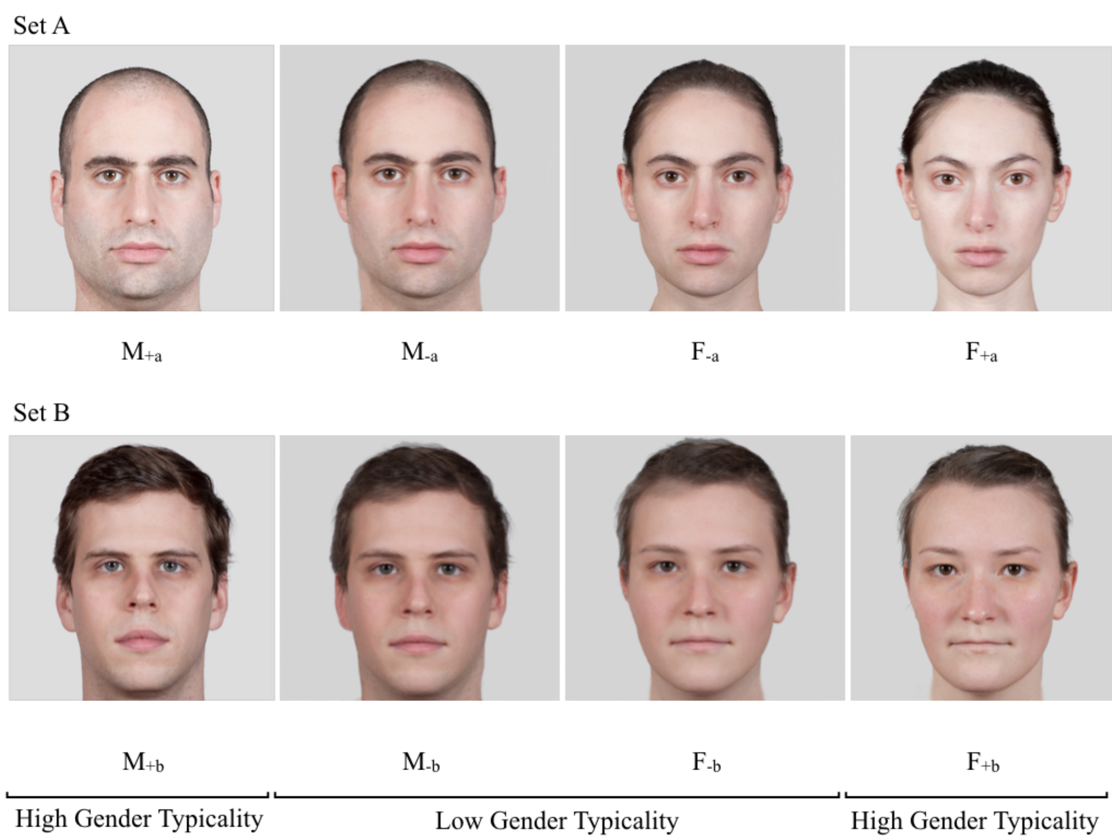

Figure 2. High gender typical faces and low gender typical morphs

Please rate the following face in terms of its gender by typing the number in the “Click to add notes” section.

1 Extremely **masculine**	2 Very **masculine**	3 A little bit **masculine**	4 Neither **masculine** nor **feminine**	5 A little bit **feminine**	6 Very **feminine**	7 Extremely **feminine**

Figure 3. Rating questionnaire of gender typicality of morphed stimuli

Please rate the following face in terms of its attractiveness by typing the number in the “Click to add notes” section.

1 Extremely **unattractive**	2 Very **unattractive**	3 A little bit **unattractive**	4 Neither **unattractive** nor **attractive**	5 A little bit **attractive**	6 Very **attractive**	7 Extremely **attractive**

Figure 4. Ratings of attractiveness of morphed stimuli

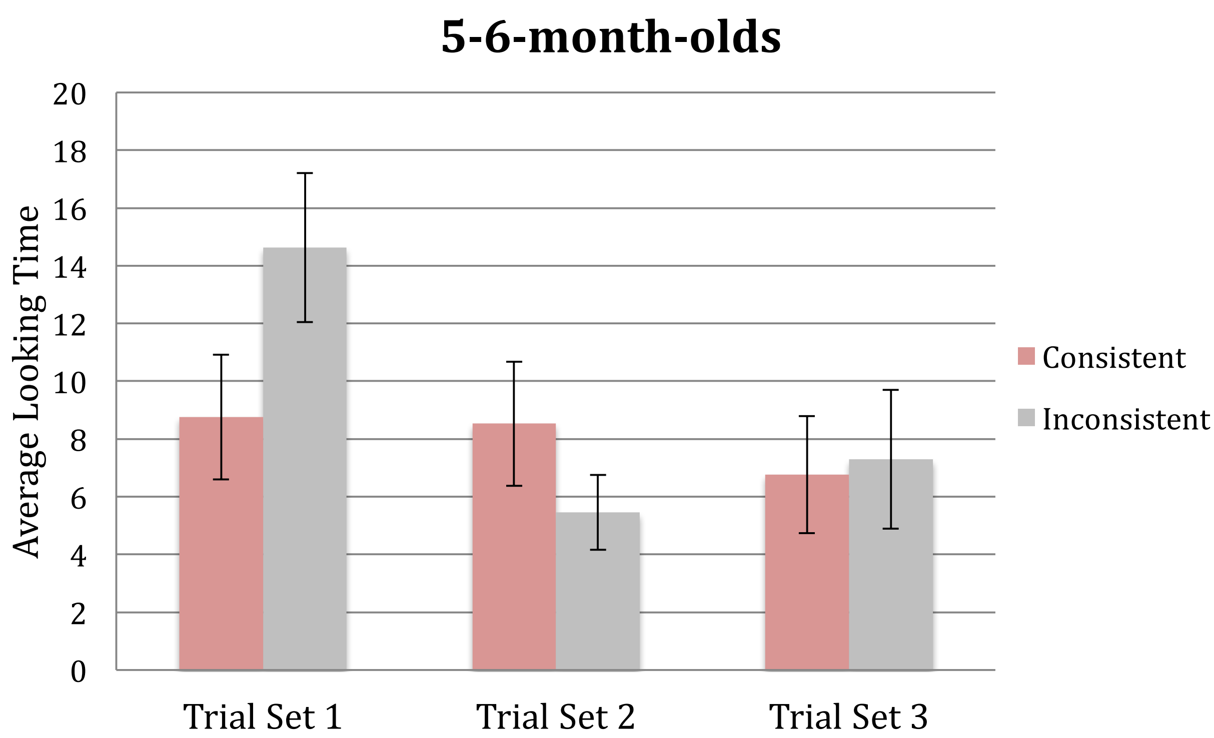

Figure 5. 5-6 mos Average Looking Time by Trial

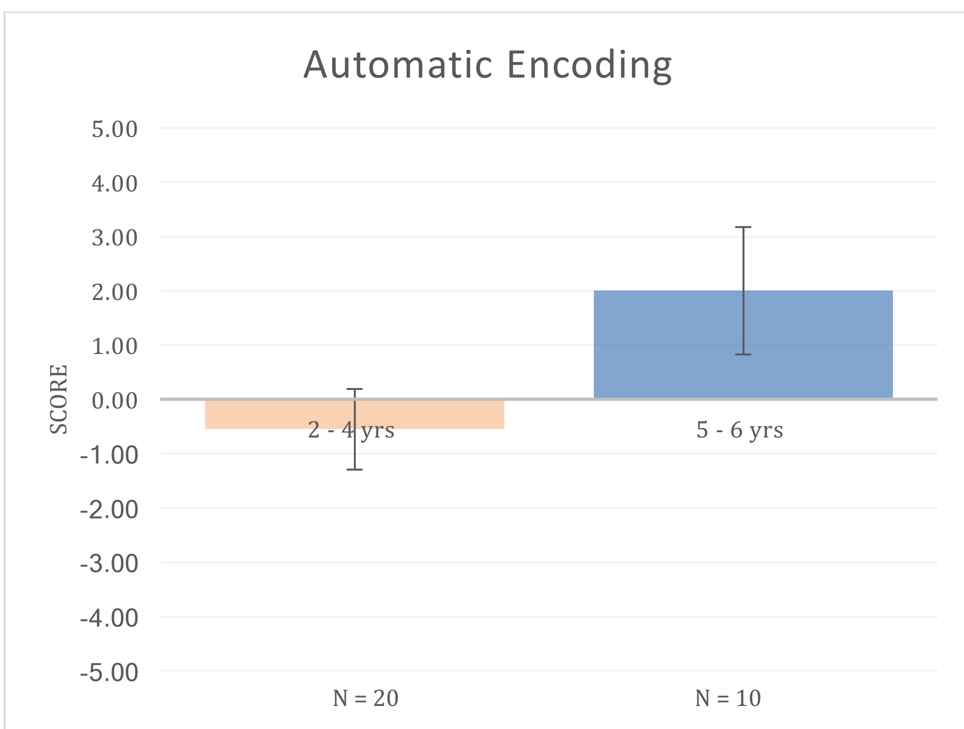

Figure 6. Gender Automatic Encoding Scores by Age

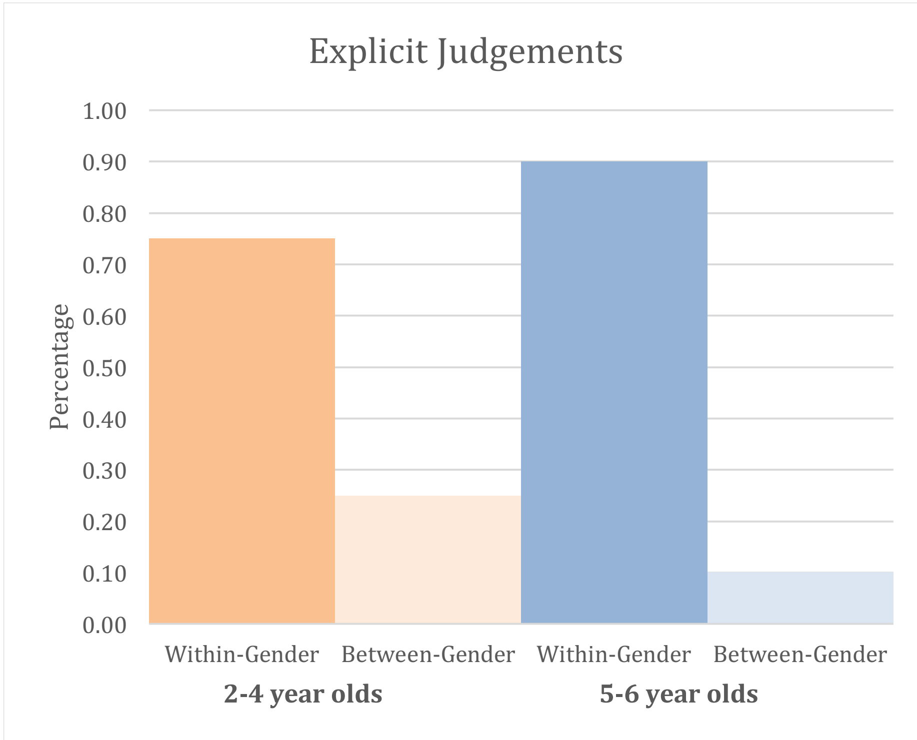

Figure 7. Explicit Judgements of Stimuli