Children and parent's looking preferences to gender-typed objects: Evidence from eye-tracking

Lauren Spinner

Lindsey Cameron

Heather J Ferguson

School of Psychology, University of Kent

Word count: 12,694

Corresponding author (Permanent address):

Dr Lauren Spinner

Tavistock & Portman NHS Foundation Trust

120 Belsize Lane

London

NW3 5BA

Email: lspinner@tavi-port.nhs.uk

Abstract

Differences between children and parents' implicit and explicit gender stereotypes were investigated in two experiments. For the first time, the visual world paradigm compared parents' and 7-year-old children's looking preferences towards masculine- and feminine-typed objects stereotypically associated with a story character's gender. In Experiment 1, participants listened to sentences that included a verb that inferred intentional action with an object (e.g. 'Lilly/Alexander will *play* with the toy"), and in Experiment 2 the verb was replaced with a neutral verb (e.g. 'Lilly/Alexander will *trip over* the toy"). A questionnaire assessed participants' explicit gender stereotype endorsement (and knowledge, Experiment 2) of children's toys. Results revealed that parents and children displayed similar implicit, but different explicit, stereotypes to one another. In Experiment 1, both displayed looking preferences towards the masculine-typed object when the story character was male, and looking preferences toward the feminine-typed object when the character was female. No gender effects were found with a neutral verb in Experiment 2, reinforcing the impact of gender stereotypes on implicit processing, and showing that the effects are not simply driven by gender stereotypic name-object associations. In the explicit measure, parents did not endorse the gender stereotypes related to toys, but appeared egalitarian, whilst children's responses were gender-stereotypic.

Keywords: gender stereotypes; children; parents; implicit attitudes

Children and Parent's Looking Preferences to Gender-Typed Toys: Evidence from Eye-Tracking

There is extensive literature showing the existence of gender-typed toy preferences in children (Weisgram, Fulcher, & Dinella, 2014; Wong & Hines, 2015), and the role that parents play as socializing agents in forming these preferences from a young age (McHale, Crouter, & Tucker, 1999; McHale, Crouter, & Whiteman, 2003; Halpern & Perry-Jenkins, 2016). However, little is known about the patterns of children and parents' implicit gender stereotypes. Parents' implicit gender stereotypes may divulge non-verbal and verbal cues, informing boys and girls about what parents believe is gender appropriate. Crucially, children predict that their parents possess gender-stereotypic attitudes in relation to toys, yet parents explicitly contradict this when asked (Freeman, 2007; Endendijk, Groeneveld, van Berkel, Hallers-Haalboom, Mesman, & Bakermans-Kranenburg, 2013). This suggests that parents may be guiding children's beliefs through other means, such as non-verbal cues or behavior. The present study examined parent's and children's implicit and explicit gender stereotypes using an established paradigm from language research, applied to this field for the first time. Specifically, we used the visual world paradigm (VWP; Cooper, 1974) to monitor participants' eye movements around visual scenes, time-locked to an auditory sentence. This paradigm enabled us to examine parents' and children's gender-related implicit biases towards gender-typed objects in real-time (i.e. as the sentence unfolded). Explicit gender stereotypes were measured using self-report questionnaires.

Children's Gender-Typed Toy Preferences

Children build their understanding of gender during the pre-school years, developing rigid definitions of how boys and girls should behave by actively seeking gendered cues from their environment (Martin & Ruble, 2004). One domain in which strict gender divisions and cues are particularly visible is children's toys (Cherney & London, 2006; Todd et al., 2017).

Toy play is a fundamental part of young children's daily experience and children frequently choose toys based on gender associations (Wood, Desmarais, & Gugula, 2002). Gender-typed toy preference is in fact one of the largest gender differences found in developmental psychology (Cherney & London, 2006), with a recent meta-analysis showing across context, geographical location, and publication date, that boys are more likely to play with masculine-typed toys than girls, and girls are more likely to play with feminine-typed toys than boys (Cohens $d = 1.03$, $p<.0001$ and Cohen's $d = -0.91$, $p<.0001$ respectively; Todd et al., 2017).

The types of toys children play with shape their cognitive and social development (Cherney, Kelly-Vance, Glover, Ruane, & Ryalls, 2003). For example, play with masculine-typed toys, such as trucks, tools, remote-control cars, and train sets (Blakemore & Centers, 2005) typically requires the use of spatial-awareness skills to manipulate the moving parts, promoting cognitive development (De Lisi & Wolford, 2002; Jirout & Newcombe, 2015). Whereas play with feminine-typed toys, such as soft animals, dolls, tea-sets, and toy kitchens (Blakemore & Centers, 2005) involves mimicking interactions, nurturance, and co-operation and empathy skills, promoting social development (Coyne, Linder, Rasmussen, Nelson, & Birkbeck, 2016; Li & Wong, 2016). If toy play is primarily in accordance with gender norms, children are potentially at a developmental disadvantage because their opportunities to develop different skills are limited.

There is research to show that exposure to gender-typed toys influences gendered beliefs and behaviors (Coyne, Linder, Rasmussen, Nelson, & Collier, 2014) as well as links between gender-typed toy play and somewhat negative outcomes in girls. For example, an experimental study by Sherman & Zurbriggen (2014) found that, in comparison to a control condition, exposing 4-7 year old girls to a Barbie doll lead to a reduction in the number of career options they believed were open to them (compared to the number of career options they believed to be open to boys). More broadly, gender stereotypes have also been shown to negatively affect young girls' perceptions of their own intellectual ability (Bian, Leslie, & Cimpian, 2017), and self-efficacy in relation to STEM (Master, Cheryan, Moscatelli, & Meltzoff, 2017). Whilst toys are a significant component of a child's learning environment, it is important to consider the role of parents here, as they are often the ones to choose which toys are purchased (Mesman & Groeneveld, 2018).

Parents as a Socializing Agent of Children's Gender Stereotypes

It is widely accepted that gender roles can be learned through social and environmental cues (Bussey & Bandura, 1999; Mischel, 1966), with social cognitive theory positioning the family as central to children's gender development (Bandura & Bussey, 2004; Tenenbaum & Leaper, 2002). During the early years of childhood, parents use their child's sex to guide expectations about their child's interests and behavior, which is communicated through gendered parenting practices (Martin, Ruble, & Szkrybalo, 2002; Mesman & Groeneveld, 2018). Parents thus create a 'gendered world' for their children by shaping their environment (Pomerleau, Bolduc, Malcuit, & Cossette, 1990) and modelling gendered behavior (Halpern & Perry-Jenkins, 2016).

There is research to show that parents rate cross-gender-typed toys as less desirable for their children than same-gender-typed or gender-neutral toys, and this is stronger among parents with more traditional gender role attitudes (Kollmayer, Schultes, Schober, Hodosi, & Spiel, 2018). Children are aware of parents' gendered expectations of play (Freeman, 2007), and these gendered expectations may influence children's play behavior and developing gender stereotypes (Kollmayer et al., 2018). For example, mixed existing literature has suggested that a small but significant correlation exists between parents' explicit gender schemas and children's explicit gender-related cognitions. From their meta-analysis of 43 studies, Tenenbaum & Leaper (2002) concluded that children were more likely to display gender-stereotypical cognitions about themselves or others if their parents possessed more traditional gender schemas.

Interestingly, Freeman (2007) showed that while children (aged 3-5 years old) expect their parents to hold stereotypical views about what boys and girls should play with, parents tended to reject common gender stereotypes regarding toys when surveyed directly (see also Campenni, 1999; Endendijk et al., 2013; Idle, Wood, & Desmarais, 1993). This contradiction raises the question of whether parents project implicit cues to their children about their gender stereotypes, which are inhibited when responding to explicit questions. This explicit response behavior may be driven by parents' social desirability or self-presentation concerns (Axinn, Young-DeMarco, & Ro, 2011; Hewstone, Rubin, & Willis, 2002; Nosek, 2005) or a lack of awareness of their own gender stereotypes (Kunda & Spencer, 2003; White & White, 2006).

Mesman and Groeneveld (2018) argue that in societies which value gender egalitarianism, there is little recent evidence to demonstrate explicit gendered parenting practices, due to a reluctance to report. However, they suggest that the majority of gender socialization practices are both specific and implicit (see also Nosek, Banaji, & Greenwald, 2002). Implicit gendered messages can be direct (e.g. responding to a particular behavior differently if it is exhibited by a boy versus a girl; see Hastings, McShane, Parker, & Ladha, 2007), or indirect (e.g. evaluating another's behavior in the child's presence). Mesman and Groeneveld (2018) posit that the sex differences which are absent in infants (Else-Quest, Hyde, Goldsmith, & Van Hulle, 2006), but become apparent during early childhood (Blakemore, Berenbaum, & Liben, 2009) are a result of primarily implicit gendered parenting. This supports the need to further examine implicit gender stereotypes in parents and children since they do not rely on conscious awareness, and are unaffected by social desirability.

Implicit Stereotypes

From the age of 10 years, children can monitor and modify their own explicit attitudes and behaviors (Fitzroy & Rutland, 2010), thus it is important to consider both explicit *and* implicit measures when predicting behavior, since implicit processing operates largely outside of conscious awareness. In the field of social development, explicit stereotypes have been extensively measured via self-report questionnaires, whereas the research that has explored implicit stereotypes has predominantly used implicit association tests (IAT; Greenwald, McGhee, & Schwartz, 1998) or the action interference paradigm (AIP; Banse, Gawronski, Rebetez, Gutt, & Morton, 2010). Including implicit measures in investigations of gender stereotypes allows researchers to assess unconscious, automatic stereotypes which are activated by the presentation of an object or person. These measures are useful for examining attitudes, particularly of young children, because they allow assessment of automatic aspects of social cognition of which children may not be consciously aware (Cvencek, Greenwald, & Meltzoff, 2011a). For example, Serbin, Poulin-Dubois, Colburne, Sen, and Eichstedt (2001) examined boys and girls aged 12, 18, and 24 months using an adapted preferential looking paradigm. The children were shown photos of vehicles or dolls, and significant preferences were found for toys stereotypically associated with children's gender at 18 months of age.

Research on the related topic of prejudice has shown that children start to display reduced *explicit* racial biases around the age of 10 years, however, *implicit* biases remain stable (Baron & Banaji, 2006; Rutland, Cameron, Milne, & McGeorge, 2005). Using the IAT, meta analyses and empirical studies have demonstrated moderate positive correlations between implicit and explicit measures in adults (e.g. Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005; Nosek, 2005), however mixed results have been found in children. In relation to gender attitudes, some studies have shown a positive correlation between implicit and explicit measures in children (Cvencek et al., 2011a). Others have found age-related declines in this correlation among males but not females, with females displaying far stronger implicit own-gender biases than males, which remain stable over time (Dunham, Baron, & Banaji, 2016). Crucially, studies that have examined the link between attitudes and actual behavior have shown that implicit measures are a better predictor of performance and behavior than explicit measures (e.g. Nock & Banaji, 2007; Nosek, Banaji, & Greenwald, 2002; Steffens, Jelenec, & Noack, 2010). Interestingly, the stereotypic effect of gender on behavior is stronger in girls than boys (Steffens et al., 2010).

There are several key studies which have examined both parents' and children's implicit and explicit gender cognitions: First, Meyer and Gelman (2016) investigated the link between parents' and 5-7-year-old children's gender essentialist beliefs using an adapted IAT related to gender-typed toys, and an explicit self-report measure of gender-stereotyping and gender-typed preferences. They found that parent's (explicit) gender essentialism predicted children's (explicit) gender-typed preferences, but not children's gender essentialism or gender-stereotypes. In other words, children of parents who assumed gender categories as natural and appropriate demonstrated more gender-stereotypical toy and activity preferences (for themselves), but there was no relation between parent's essentialism and their children's beliefs about what is appropriate for others (in relation to gender-typed occupations). Second, Endendijk et al. (2013) examined parents' and 3-year-old children's implicit gender stereotypes using the AIP (parents and children) and the IAT (parents only), and explicit stereotypes using a self-report questionnaire (parents only). They found that girls' implicit gender stereotypes were strongly related to their mother's implicit gender stereotypes, but the same correlation was not evident amongst boys and their mothers. More recently, a study examining relationships between Chilean parents' and their children's math-gender stereotypes found that mothers and fathers associated math with boys in both implicit and explicit measures, boys implicitly showed this stereotype more strongly than girls, and girls' implicit math self-concept was partially explained by parental beliefs. This demonstrates that by 5 years of age, before starting school, parental beliefs are contributing to the development of children's math-gender stereotypes and self-concepts (del Río, Strasser, Cvencek, Susperreguy, & Meltzoff, 2018).

Although measures such as the IAT have been widely used to examine implicit biases, they have not been without their critique. For example, a 2013 meta-analysis found weak correlations between IAT scores and explicit ethnic and race bias in relation to both behavior and attitudes, questioning its construct validity (Oswald, Mitchell, Blanton, Jaccard, & Tetlock, 2013). The present study fills an important gap in the literature by introducing a robust and precise measure of unconscious processing, in a novel way, via the visual world paradigm. The current study also utilizes the same explicit and implicit measures with parents *and* their children, which allows for a comparison of the groups' gender-related biases.

Visual World Paradigm

In the visual world paradigm (e.g. Cooper, 1974; Tanenhaus Spivey-Knowlton, Eberhard, & Sedivy, 1995), participants' eye movements around visual scenes are recorded and time-locked to a related auditory input. For example, the sentence, "The girl will ride the carousel" could be paired with a scene containing a girl, and two 'ride-able' objects (a carousel and a motorbike; Kamide, Altmann, & Haywood, 2003). Studies of this kind show that participants incorporate cues from syntax, semantics and world knowledge to constrain the available set of objects, and move their eyes to the appropriate visual object *before* it has been mentioned in the audio. The paradigm therefore provides an implicit measure of expectation in real-time (i.e. as the sentence unfolds). The visual world paradigm has been used extensively in the field of psycholinguistics, thus its validity as a measure of real-time expectations is well established (e.g. Altmann & Kamide, 2007, 2009; Kamide, Lindsay, Scheepers, & Kukona, 2016). Moreover, while the topics of stereotypes or gender have never been examined using this approach, the paradigm has provided valuable new insights into other social psychological phenomena, including mental state inferences (e.g. Ferguson & Breheny, 2011; 2012; Ferguson, Scheepers, & Sanford, 2010), just-world theory (e.g. Callan, Ferguson, & Bindemann, 2013), and social referencing (e.g. Crosby, Monin, & Richardson, 2008).

The present study adapted the visual world paradigm to investigate children and parent's eye movements towards gender-typed objects in relation to male and female characters. This approach offers several advantages. First, as fixations are measured in high temporal resolution, this method provides precision and reveals unconscious processing that participants are not likely to be aware of. This reduces risks of demand characteristics as it would be very difficult for participants to use explicit knowledge to change their implicit biases (Berends, Brouwer, & Sprenger, 2016). Second, we developed semi-realistic visual scenes which contained a range of gender-typed objects which should be familiar to participants; this offers some assurance of external validity. Third, the 'look and listen' visual world paradigm makes low demands on executive skills, thus can be used with a wide range of participant ages. Finally, implicit measures have been shown to provide better predictive validity than explicit self-report measures of stereotyping behaviors (see Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Therefore, using the visual world paradigm allows us to gain a better understanding of the implicit biases that might underlie gender-relevant cognition among both parent and child populations; extending the limited literature in this area in important ways.

Experiment 1

Experiment 1 addressed two key research questions; first, do participants present gender-stereotypic responses in relation to children's toys/objects/activities and are these evident in both implicit and explicit measures? And second, do parents and children present similar or different gender-stereotypic responses to one another in relation to children's toys/objects/activities, and is this evident in both implicit and explicit measures? Given the gender differences seen in the reviewed research of implicit gender stereotyping (e.g. Endendijk et al., 2013; del Río et al., 2018), we were also interested to see if there were participant-gender differences in responses in the present study.

Participants were tested on the implicit visual world task and the explicit gender stereotype endorsement task. We specifically tested children aged 7 years old because gender is a highly salient social category at this age; gender-related knowledge, stereotypes, and behavior are frequently observed, but it is around this age that children also demonstrate some flexibility in their views of gender norms (Martin & Ruble, 2004). This can result in individual differences in gender-related cognitions and behavior, which makes this an interesting age group to in which to examine both implicit and explicit processes. This age group may also face social desirability pressures when presented with explicit measures, so implicit measures are a useful way to overcome these and have been utilized by other researchers when examining gender stereotypes in this age group (see Cvencek et al., 2011a; Cvencek, Meltzoff, & Greenwald, 2011b; Cvencek, Meltzoff, & Kapur, 2014; & Cvencek, Kapur, & Meltzoff, 2015).

In line with previous research, we predicted that children would demonstrate gender-stereotypic responses in their explicit gender-stereotype endorsements (i.e. that only boys should play with masculine-typed toys, and only girls should play with feminine-typed toys), but parents would show more egalitarian attitudes (i.e. indicating that both masculine and feminine-typed toys should be played with by both boys and girls). In contrast, in the implicit gender stereotype measure we predicted that both children and parents would show gender-typed biases in their looking preferences (i.e. looking towards masculine-typed objects for a male story character, and towards feminine-typed objects for a female story character). Thus, we expected children's implicit and explicit gender stereotypes to follow similar patterns, but parents' implicit and explicit gender stereotypes to diverge. Finally, as an exploratory question, we also examined whether looking preferences were biased to objects that were stereotypically associated with participants' own gender, and whether this differed between parents and children.

Method

Participants

Thirty-three British children (19 males and 14 females) aged between 7 – 8 years (mean age = 87.97 months, *SD* = 4.46), and 35 of their parents (6 males and 29 females) took part. A post-hoc analysis of power (calculated using the *simr* package in R; Green & MacLeod, 2016) showed that this sample size gives 73% power to detect the 3-way interaction between Participant Gender, Participant Age and Character Gender with the significance level of α=.05 in the eye-tracking task. Twenty-four parents provided their date of birth; their mean age was 40.67 years (*SD* = 70.08). Children were recruited from a university database. Ethical consent was gained from the university Research Ethics Committee, in line with British Psychological Society guidelines. Parental and participant consent were obtained prior to the study commencing.

Materials and Design

Gender stereotype endorsement. Based on a measure used by Weisgram, Fulcher, & Dinella (2014) and Spinner, Cameron, & Calogero (2018); participants were shown eight pictures of toys in a randomized order (four stereotypically masculine, four stereotypically feminine; from Blakemore & Centers, 2005) and asked "Who should play with this toy?". The question was phrased in this way to ensure we tapped into attitudes, i.e. endorsement of gender stereotypes, rather than gender stereotype knowledge. This follows Liben & Bigler's (2002) recommendation (see also Signorella, Bigler, & Liben, 1993). There were three response options; "Only boys" (coded as 0), "Only girls" (coded as 0), or "Both boys and girls" (coded as 1). Participants were asked to select one option by placing a tick in the column which represented their answer. Thus, there was one within-participant independent variable with two levels (Toy Type: masculine vs feminine). This task provided an explicit measure of the strength of participants' gender stereotype views in relation to children's toys. A 'flexibility score' was calculated by summing the assigned codes across the eight toys. Scores could range from 0 to 8, with higher scores indicating more gender flexible attitudes toward toy-play. It should be noted that no children or parents indicated a counter-stereotypical endorsement, for example "only boys should play with dolls". This means that all responses coded as '0' were stereotypical responses.

Audiovisual scenes. Sixteen experimental visual scenes were constructed in Photoshop using objects taken from a Google image search, and were presented on a 17-inch color monitor in 1024 x 768 pixels resolution. Each visual scene could be paired with one of two auditory sentences that described a character performing an activity or interacting with an object (e.g. "[Character] has painted a picture"); one included a female character's name and one included a male character's name. Note that the target object was deliberately described using gender neutral nouns (e.g. "toy", "picture", "costume") to avoid bottom-up effects from the language. Each visual scene included a stereotypically feminine or masculine variant of the object described in the audio sentence, among several background and distracter items (see Appendix A for all experimental scenes and accompanying audio sentences). For example, accompanying the auditory sentence, "Lilly/Alexander will play with the toy", there was a visual scene depicting a garden containing two 'playable' objects – a doll and a truck, alongside other background items including decking and grass, a tree, a watering can, and a butterfly (see Figure 1). The masculine- and feminine-typed objects were positioned in varying locations around each scene to prevent systematic viewing strategies, and were derived from previous research on masculine and feminine toys, for example Blakemore and Centers (2005). Some objects also differed in color (pink and blue), a cue typically used by children use as an indicator of whether a toy is masculine or feminine (Paoletti, 2012). Gender stereotypes for each object was verified post-hoc, with 60 students from the University of Kent (none of whom participated in the main study) using an online questionnaire platform (Qualtrics). Participants were shown 32 images of objects (16 feminine- and 16 masculine-typed), cropped from the larger visual scenes, and used a horizontal slider to indicate the gender stereotypically associated with the pictured object (scale ranging from 0 (stereotypically female) to 100 (stereotypically male). Results confirmed that all that feminine-typed objects were more strongly associated with females (i.e. all had mean scores below 50, with an overall mean rating of 14.8 across all 16 feminine-typed objects), and masculine-typed objects were more strongly associated with males (i.e. all had mean scores above 50, with an overall mean rating of 74.9 across all 16 masculine-typed objects).

One version of each item was assigned to one of two counterbalanced presentation lists, with each list containing sixteen experimental items, eight describing a female character and eight describing a male character. In addition, 16 filler scenes were interspersed randomly among the sixteen experimental items to create a single random order. These filler items were included to distract participants from the true aim of the study. These filler scenes contained ‘neutral’ items which were not related to children’s toys or activities (see Appendix B for all filler scenes and accompanying audio sentences). For example, one scene depicted a doctor’s office, paired with the auditory sentence “Doctor Davis will put on the jacket”, where two jackets (a white jacket and a sports jacket) were pictured. Half of the participants saw list 1 and the other half saw list 2; children and their parents always viewed the same list.

Figure 1

Example of an experimental visual scene.

Thus, the audiovisual task employed a 2 (Participant Gender: male vs female) x 2 (Participant Age: adult vs child) x 2 (Character Gender: male vs female) mixed design, with Participant Gender and Age as the between-participants variables, and Character Gender as the within-participants variable. The dependent variable was the proportion of time-bins within each trial in which participants fixated on the critical objects (i.e. the masculine/feminine items).

Procedure

Eye tracking was recorded from the participants' dominant eye using an EyeLink 1000 eye-tracker (viewing was binocular), running at 1000Hz. The experiment was controlled using Experiment Builder software. At the beginning of the experiment, and at the half-way point (or as needed), the eye-tracker was calibrated and validated against nine fixation points, using the standard EyeLink calibration procedures. Participants were told that they would see images on the computer screen and these would be accompanied by a spoken sentence, presented through the loudspeaker. They were informed that their task was to listen to the sentences whilst simultaneously viewing the accompanying visual scenes. Each trial began with a centrally-located drift correction procedure, followed by the target image, along with the auditory sentence. The onset of the image preceded the onset of the corresponding sentence by 1000ms. The picture stayed on-screen for the duration of the sentence (approximately 2000ms), with the auditory sentence typically ending 4000ms before the end of the trial. This part of the experiment took approximately 10 minutes to complete.

Participants then completed the explicit gender stereotype endorsement measure (researcher-guided for children) and were debriefed by the experimenter. The entire experiment took approximately 20-25 minutes to complete, and the child was always tested first, then the parent.

Results

The full datasets and analysis scripts are available on the Open Science Framework web pages (https://osf.io/24b93/).

Explicit Measure

Gender stereotype endorsement of toys. To examine participants' explicit gender stereotype endorsement, a one-way ANOVA was performed on flexibility scores, with Participant Gender and Participant Age included as between-subjects factors. Analyses revealed a main effect of Participant Age, $F(1, 61) = 82.88$, $p < .001$, $\eta_p^2 = .58$, whereby parents achieved significantly higher gender flexibility scores ($M = 7.03$, $SD = 1.81$) than children ($M = 1.63$, $SD = 1.62$), Supporting hypotheses, parents assigned more toys to the "both boys and girls" category than children did, indicating that parents are more flexible about who should play with the toys than children. There was no effect of Participant Gender ($p = .412$; Males: $M = 3.93$, $SD = 1.60$; Females: $M = 4.39$, $SD = 1.96$), and the interaction between Participant Age and Participant Gender was also non-significant ($p = .445$) [1].

Implicit Measure – Eye-Tracking

Eye-tracking data processing. Eye movements initiated while the target image was onscreen were processed according to the relevant picture and sound onsets. The spatial coordinates, in pixels, of fixations were mapped onto the appropriate regions of analysis, corresponding to the masculine and feminine objects in each image. If a fixation was located within 20 pixels of a visual object's perimeter, it was coded as belonging to that object, otherwise, it was coded as background.

Visual preferences to these two objects were examined by calculating a masculine/feminine-object advantage score as a function of time (i.e. the probability of fixating on the 'feminine' object *minus* the probability of fixating on the 'masculine' object). This measure is therefore symmetrical around zero such that higher proportions of fixations on the "feminine" object result in a positive score, whereas higher proportions of fixations on the "masculine" object result in a negative score.

Statistical analyses were carried out on log-transformed masculine/feminine advantage scores using linear mixed-effects models, as described below. Models included the maximal random effects structure, including random effects for participants and items (which allows generalization of significant effects across participants and items), and crossed random

[1]Note: A between-subjects, rather than a fully-crossed within-subjects, design was used for analysis of both implicit and explicit data because it was not possible to pair parents with their children specifically in every instance: Sometimes both parents took part and sometimes more than one child took part (e.g. twins).

and crossed random slopes for each of the independent variables (as suggested by Barr, Levy, Scheepers, & Tily, 2013). Random effects were only removed where they lead to non-convergence due to overparameterization. A significance level of 5% was used for all tests.

Preview analysis. Onset of the picture preceded the onset of the auditory sentence by 1000ms; thus, we first analyzed the distribution of fixations on masculine/feminine objects during this preview period to examine whether any very early visual biases emerged before the onset of the auditory sentence. That is, this preview analysis allowed us to test whether participants' own gender preferences influenced their initial visual biases to objects in the scene. These data are plotted in Figure 2. Analyses compared masculine/feminine advantage scores between 500-1000ms post-picture onset (since prior to this eye movements are still being initiated; see Ferguson & Breheny, 2011; Ferguson et al., 2010). Each model included the independent variables of Participant Gender and Participant Age as fixed effects, which were deviation coded (-.5 *vs.* .5) to ensure they could be directly compared.

Results revealed only a trend for different object preferences between female and male participants (*Est.* = -.468, $t = -1.88$, $p = .075$), with females showing a preference to fixate on the feminine objects ($M = 0.85$) and males showing little preference between feminine and masculine objects ($M = - 0.02$). There was no effect of Participant Age, or a significant interaction (ps > .23).

Main analyses. The main analyses focused on visual biases during the auditory sentence. To examine this, temporal onsets and offsets of the fixations were recalculated on a trial-by-trial basis, relative to the onset of the verb (e.g. "play") in the corresponding auditory input (since this is the earliest point a target object can be accurately inferred). A time period ranging from 1000ms before the onset of this verb to 3000ms after the onset of the verb was examined, since it includes the average character name onset ($M = -761$ms; range -1157 to -

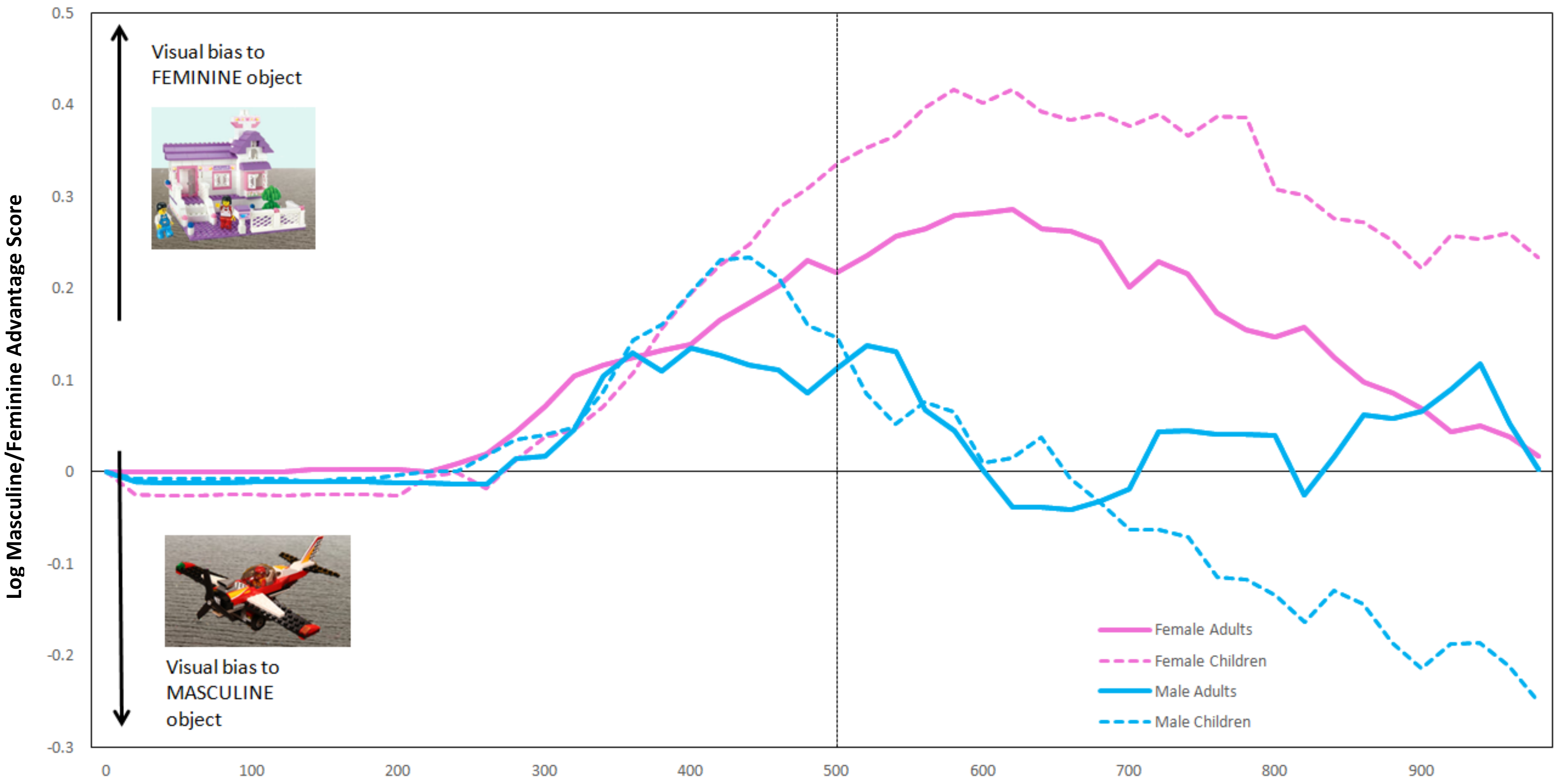
620ms) and allows sufficient time beyond the neutral object (e.g. "toy") to examine changes in visual biases. Figures 3 and 4 plot the average masculine/feminine advantage scores in each condition, for every 20ms time-slot within the selected time period. The solid black line in both figures (t=0) indicates the onset of the verb, and arrows/labels show the average onset of other relevant auditory information.

In order to reduce the number of statistical tests without masking potentially important detail, the 20ms data points were aggregated into larger analysis windows; specifically, we calculated eight consecutive 500ms time windows. We note that visual world eye-tracking data can be aggregated and analyzed in numerous ways. For example, the time periods for analysis can be defined according to word onsets and offsets (e.g. Altmann & Kamide, 1999; Ferguson & Breheny, 2011), fixed length time windows (e.g. Callan et al., 2013; Mirkovic & Altmann, 2019), a cluster-analytic procedure (e.g. Ferguson et al., 2010), or more recently using growth-curve or permutation analysis to model condition effects over time (e.g. Black, Barzy, Williams, & Ferguson, 2019). We chose to use fixed length time windows since the language input did not provide consistent markers for time-locking, and we had no clear predictions on exactly *when* condition effects would emerge.

Thus, this approach allowed an unbiased exploration of visual biases in children and their parents as the sentence unfolded, replicating the approach used in previous research that has examined social inferences (Callan et al., 2013). These eight consecutive 500ms time bins are indicated by the dashed vertical lines on Figures 3 and 4. Statistical models included the independent variables of Participant Gender, Participant Age and Character Gender as fixed effects, which were deviation coded (-.5 *vs.* .5) to ensure they could be directly compared. Statistical tests were not corrected for multiple comparisons (i.e. the eight time windows). Instead, we focused on effects that persisted across multiple consecutive time windows that would indicate a robust pattern of bias influencing eye movement data, rather than simply a

Figure 2

The average location advantage scores for each participant group during the preview period in Experiment 1.



Note. The dashed vertical line indicates the 500ms point; the timeslot following this point (500 – 1000ms) was used for statistical analysis.

Type I error. This approach was appropriate given the variability in linguistic content and word onsets/offsets in the experimental sentences, and since we did not have predictions on the exact timing of effects. The resulting statistical effects are reported in full in Table 1. Statistical analyses revealed a significant effect of Character Gender from the onset of the verb onwards.

Overall, participants showed a preference to fixate the feminine-typed objects in the female-character condition, and a preference to fixate the masculine-typed objects in the male-character condition, as hypothesized. This suggests that participants rapidly inferred male/female gender information from the character's name, and used this to direct their visual search of objects in the scene according to stereotypical knowledge of gender preferences.

However, Character Gender also interacted with other variables in some time windows, which tells us that the participants' own age and gender modulated the effect of the character's gender. Given the variability in linguistic content and word onsets/offsets in the experimental sentences used here, we focus on those effects that persisted across multiple consecutive time windows, since this suggests a robust pattern of bias is influencing the eye movement data, and is less likely to be a false positive. No significant effects were found in time window 1 (-1000 to -500ms before verb-onset), however, in time window 2 (-500 to 0ms before verb-onset) the effect of Participant Gender was significant, reflecting a looking-preference for the feminine-typed objects among female participants, and a preference for the masculine-typed objects among male participants. This pattern mirrors that seen in the preview period. This effect of Participant Gender continued to be significant into time windows 3 and 4 (0 – 1000ms from verb-onset), reflecting the same pattern of looking preferences as the previous time window. In addition, time windows 3 and 4 showed a significant 3-way interaction between Character Gender x Participant Age x Participant Gender.

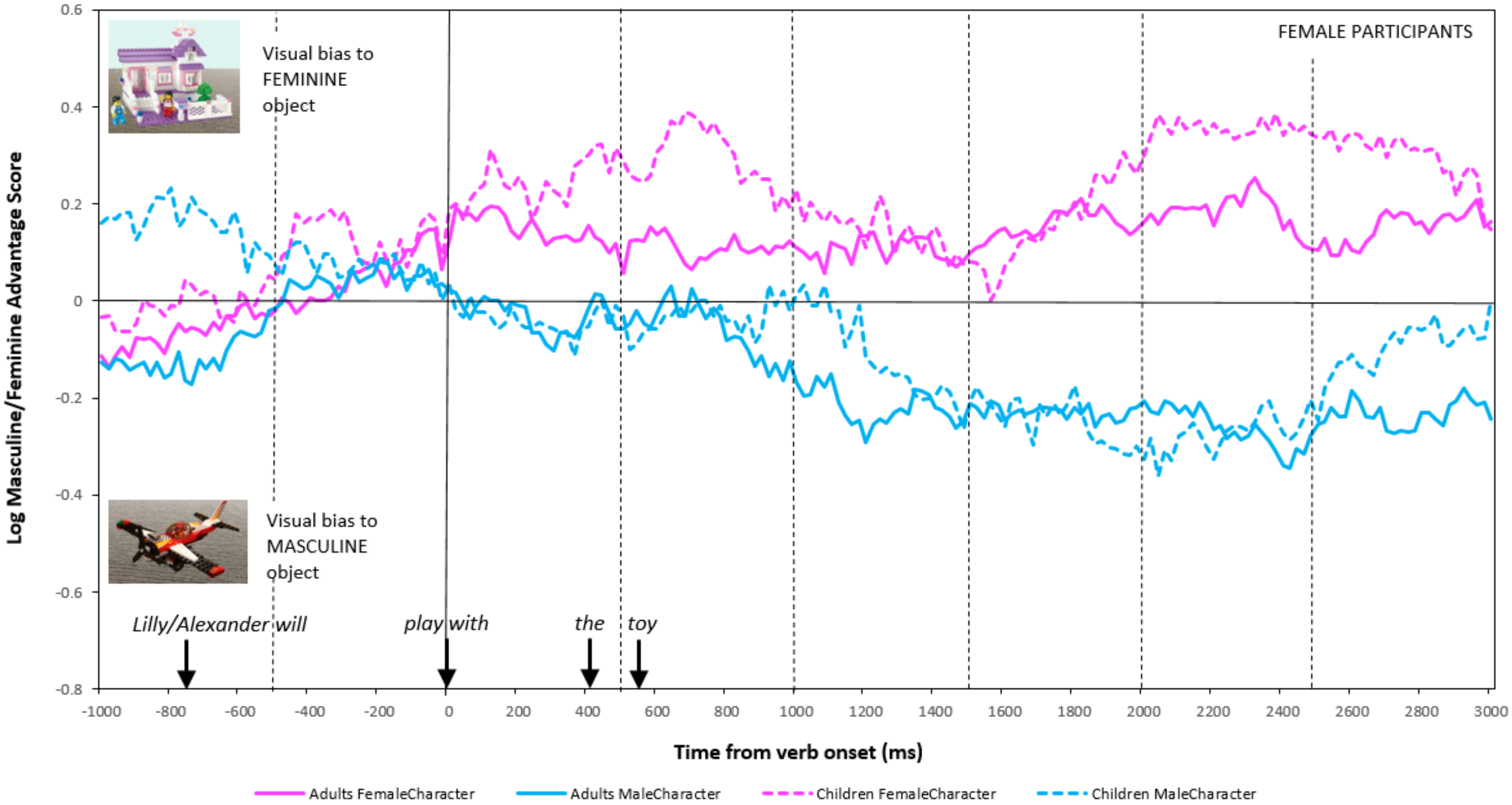
Further analyses revealed a Character Gender x Participant Age interaction among the male participants (0 to 500ms: *Est.* = 1.28, *SE* = 0.63, $t = 2.02$, $p < 0.05$; 500 to 1000ms: *Est.* = 1.34, *SE* = 0.7, $t = 1.92$, $p = 0.06$) that was not significant among the female participants (0 to 500ms: *Est.* = -0.44, *SE* = 0.58, $t = -0.77$, $p = 0.45$; 500 to 1000ms: *Est.* = -0.71, *SE* = 0.65, $t = 1.09$, $p = 0.28$). Specifically, adult males showed a preference to fixate on the feminine-typed items in the female-character trials, and the masculine-typed items in the male-character trials (0 to 500ms: *Est.* = -1.49, *SE* = 0.66, $t = -2.25$, $p = 0.06$; 500 to 1000ms: *Est.* = -1.75, *SE* = 0.56, $t = -3.14$, $p < 0.01$), but male children did not differ in their looking preferences according to Character Gender (0 to 500ms: *Est.* = -0.15, *SE* = 0.34, $t = -0.44$, $p = 0.67$; 500 to 1000ms: *Est.* = -0.46, *SE* = 0.36, $t = -1.26$, $p = 0.22$).

In time window 5 (1000 to 1500ms after verb-onset), Character Gender significantly interacted with Participant Age, showing that the effect of Character Gender (i.e. a feminine object bias for female characters and a masculine object bias for male characters) was larger among adults (*Est.* = -1.39, *SE* = 0.39, $t = -3.59$, $p = 0.001$) than children (*Est.* = -0.78, *SE* = 0.25, $t = -3.13$, $p = 0.002$). This pattern persisted into time window 6 (1500 to 2000ms after verb-onset), but here it was subsumed under a significant 3-way interaction between Character Gender x Participant Age x Participant Gender, which lasted throughout the following two time windows (until 3000ms after verb-onset).

Further analyses in these three time windows revealed the same pattern as seen in time windows 3 and 4 (0-1000ms), with a Character Gender x Participant Age interaction that was significant among the male participants (1500 to 2000ms: *Est.* = 2.87, *SE* = 0.73, $t = 3.93$, $p < 0.001$; 2000 to 2500ms: *Est.* = 2.94, *SE* = 0.99, $t = 2.95$, $p < 0.01$; 2500 to 3000ms: *Est.* = 2.57, *SE* = 0.66, $t = 3.87$, $p < 0.001$) but not among the female participants (1500 to 2000ms: *Est.* < 0.01, *SE* = 0.66, $t = -0.01$, $p = 0.99$; 2000 to 2500ms: *Est.* = -0.8, *SE* = 0.86, $t = -0.93$, $p = 0.36$; 2500 to 3000ms: *Est.* = -0.2, *SE* = 0.84, $t = -0.24$, $p = 0.81$).

Figure 3

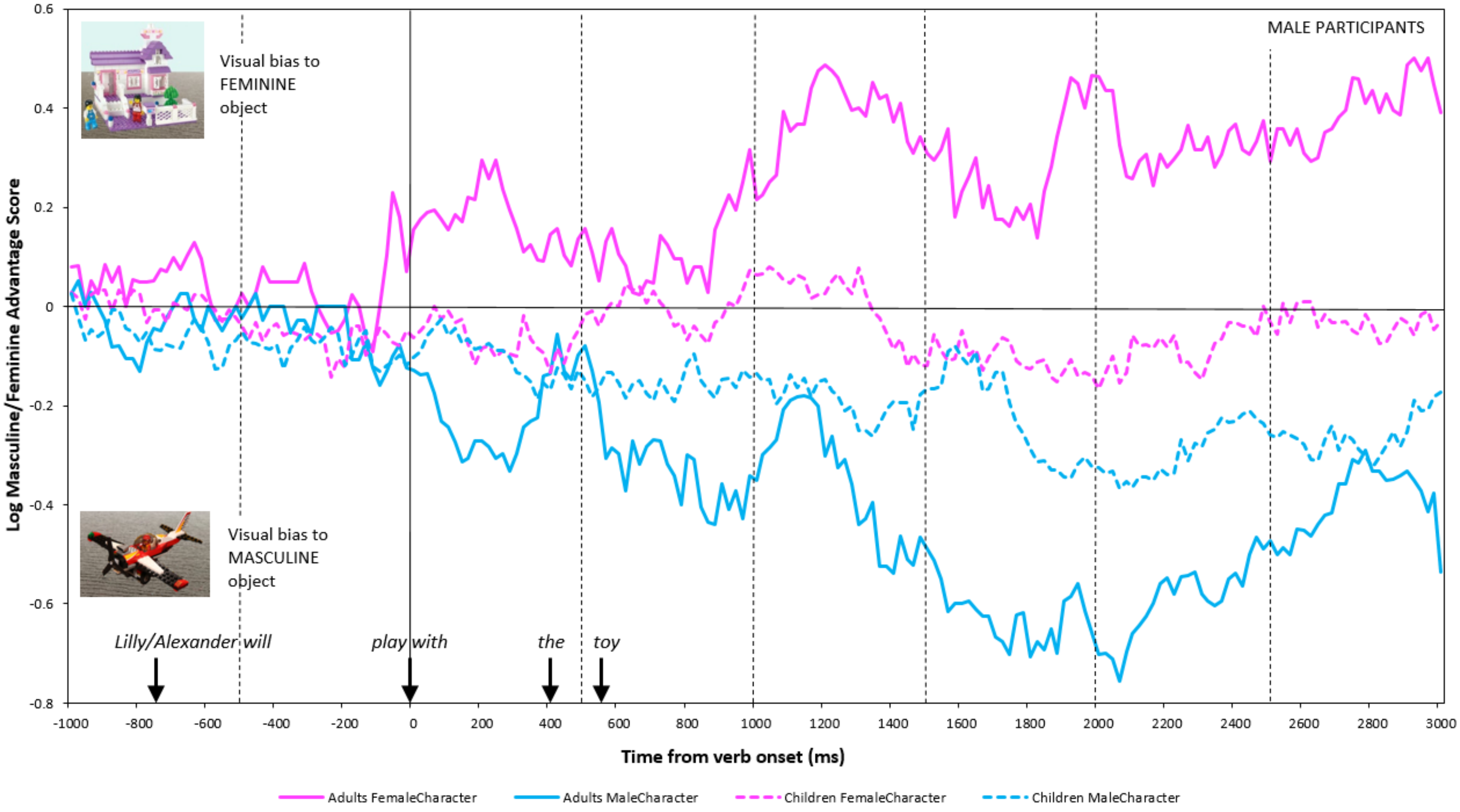
The average location advantage scores for female participants for each condition in Experiment 1.



Note. The solid black vertical line (t=0) indicates the onset of the verb in the audio sentence. The dashed vertical lines represent the 500ms timeslots that were used for statistical analysis.

Figure 4

The average location advantage scores for male participants for each condition in Experiment 1.



Note. The solid black vertical line (t=0) indicates the onset of the verb in the audio sentence. The dashed vertical lines represent the 500ms timeslots that were used for statistical analysis.

Table 1

*Results of Linear Mixed Effects Models for each time window of interest in Experiment 1, where * p<.05, ** p<.01, *** p<.001.*

		Character Gender	Participant Age	Participant Gender	Character Gender x Participant Age	Character Gender x Participant Gender	Participant Age x Participant Gender	Character Gender x Participant Age x Participant Gender
	Est.	-0.03	0.24	0.10	0.47	-0.35	-0.73	-0.69
-1000 to -500ms	*SE*	0.19	0.22	0.19	0.38	0.38	0.37	0.76
	t	0.16	1.08	0.53	1.24	-0.92	-1.95	-0.90
	Est.	-0.29	0.05	-0.44	0.09	-0.18	-0.5	0.92
-500 to 0ms	*SE*	0.20	0.18	0.21	0.37	0.37	0.37	0.74
	t	-1.44	0.26	-2.13*	0.24	-0.49	-1.38	1.24
	Est	-0.84	-0.06	-0.51	0.46	-0.02	-0.35	1.89
0 to 500ms	*SE*	0.24	0.22	0.19	0.46	0.46	0.37	0.91
	t	-3.55***	-0.27	-2.74**	1.00	-0.05	-0.94	2.07*
	Est.	-1.03	0.23	-0.59	0.20	-0.02	-0.13	2.04
500 to 1000ms	*SE*	0.29	0.26	0.21	0.51	0.50	0.39	1.01
	t	-3.56**	0.88	-2.82**	0.39	-0.04	-0.33	2.02*
	Est.	-1.38	-0.03	-0.08	1.20	-0.78	-0.66	1.86
1000 to 1500ms	*SE*	0.30	0.21	0.22	0.51	0.50	0.41	1.01
	t	-4.60***	-0.14	-0.37	2.35*	-1.56	-1.61	1.84
	Est.	-1.72	0.01	-0.40	1.53	-0.48	0.13	3.11
1500 to 2000ms	*SE*	0.34	0.23	0.23	0.53	0.52	0.42	1.05
	t	-5.05***	0.06	-1.74	2.87**	-0.91	0.31	2.96**
	Est.	-2.16	0.06	-0.58	1.20	-0.23	-0.29	4.16
2000 to 2500ms	*SE*	0.38	0.21	0.19	0.69	0.66	0.38	1.32
	t	-5.63***	0.28	-3.05**	1.73	-0.34	-0.76	3.15**
	Est.	-1.82	<0.01	-0.42	1.26	-0.55	-0.93	2.91
2500 to 3000ms	*SE*	0.34	0.22	0.24	0.65	0.63	0.42	1.25
	t	-5.31***	<0.01	-1.74	1.94	-0.87	-2.22*	2.32*

Note: Data reported here are uncorrected for multiple comparisons. See Experiment 1 “Main analyses” section for rationale.

As before, adult males showed a significant effect of Character Gender in all time windows (1500 to 2000ms: *Est.* = -3.32, *SE* = 0.65, t = -5.12, $p < 0.005$; 2000 to 2500ms: *Est.* = -3.63, *SE* = 0.78, t = -4.64, $p < 0.005$; 2500 to 3000ms: *Est.* = -3.25, *SE* = 0.59, t = -5.49, $p < 0.005$), but this effect was non-significant or reduced among the male children (1500 to 2000ms: *Est.* = -0.45, *SE* = 0.37, t = -1.22, p = 0.24; 2000 to 2500ms: *Est.* = -0.71, *SE* = 0.33, t = -2.17, p = 0.03 ; 2500 to 3000ms: *Est.* = -0.72, *SE* = 0.41, t = -1.75, p = 0.1). In addition, time window 7 (2000 to 2500ms after verb-onset) showed an overall effect of Participant Gender, similar to that seen in time windows 2 to 4, where female participants preferentially fixated the feminine-typed objects and male participants preferentially fixated the masculine-typed objects[2].

Discussion

Experiment 1 revealed a dissociation between explicit and implicit measures in parents but not children: Parents showed greater gender flexibility than children on the explicit toy-rating measure, but comparably strong implicit gender stereotypes on the implicit eye-tracking measure. That is, when explicitly questioned, parents were more likely to assign toys to the "both boys and girls" category than children, who adhered to gender stereotype norms (i.e. they assigned masculine toys to boys and feminine toys to girls). In contrast, when eye-tracking measured implicit gender biases, both parents and their children showed gender-stereotyped looking behavior (i.e. a preference to fixate the feminine-typed objects in the female-character condition, and a preference to fixate the masculine-typed objects in the male-character condition). Crucially, this bias emerged incrementally as the sentence progressed; it was evident immediately following the verb (e.g. "*play"*). This shows that participants made rapid inferences about the intended object, and biased their domain of reference according to stereotypes based on the character's gender.

[2]Note: It was not possible to correlate the explicit and implicit data due ceiling effects, low variance, and negative skewness in the parents' explicit gender stereotypes scores in both Experiments 1 and 2.

This interpretation of the implicit eye movement data assumes a strong influence from top-down gender stereotypes for participants to move their eyes to predict the character's actions in the sentence. However, it is possible that this pattern reflects simpler bottom-up processing, where informed knowledge about gender associations drive eye movements based on semantic associations between the character's name and gendered objects (e.g. we more frequently encounter girls playing with dolls than cars in everyday life), rather than more socially-situated gender stereotypes. We tested this possibility in Experiment 2.

Experiment 2

Experiment 2 replicated the design from Experiment 1, but modified the auditory sentences to include a neutral verb (e.g. "*Jack will trip over the toy*"). Here, while learned gender associations based on the character's name to specific toys should still be present, the verb now means that the character's action on the target object is gender neutral - the two toys are equally likely to be tripped over. Thus, if the effects seen in Experiment 1 are simply due low-level semantic associations based on familiar gender-object combinations, then effects in Experiment 2 should replicate those in Experiment 1 (i.e. visual preferences to objects based on character gender). However, if the effects are driven by socially-derived gender-stereotypes, we predicted that gender-based biases to masculine/feminine objects would be eliminated with the neutral verb in Experiment 2. In addition, to distinguish between gender-stereotype knowledge and attitudes (as previous research has documented; see Banse et al., 2010), we adapted the explicit measure in Experiment 2 to assess both gender-stereotype endorsement (as in Experiment 1: "*Who should play with this toy?*") and gender-stereotype knowledge ("*Who plays with this toy?*").

All methodological and analysis procedures for Experiment 2 were pre-registered on the Open Science Framework (OSF) web pages (see https://osf.io/24b93/).

Method

Participants

Thirty-three British children (12 males and 21 females) aged between 7 – 8 years (mean age = 88.24 months, *SD* = 3.54), and 33 of their parents (4 males and 29 females) took part. Twenty-eight parents provided their date of birth; their mean age was 38.25 years (*SD* = 68.96). This sample was chosen to closely match that used in Experiment 1.

Materials and Design

Gender stereotype endorsement and knowledge. The materials and administration of this measure were the same as Experiment 1, except that an additional question was asked after each toy-picture (*"Who should play with this toy?*" and "*Who plays with this toy?").* The order in which these questions were presented was counterbalanced, with children and their parents always receiving the questions in the same order as one another. An endorsement score and a knowledge score was calculated separately by summing the assigned codes across the eight toys. Scores could range from 0 to 8 for each variable, with higher scores indicating greater gender flexibility, as in Experiment 1. No children or parents indicated a counter-stereotypical endorsement, for example "only boys should play with dolls". This means that all responses coded as '0' were stereotypical responses.

Audiovisual scenes. The same 16 experimental visual scenes from Experiment 1 were included in Experiment 2, but here the verb in auditory sentences was changed to a neutral verb with no gender stereotype connotations (e.g. "*[Character] has seen a picture*"; (see Appendix A).

All other aspects of the experimental design and procedure remained the same as Experiment 1.

Results

All analysis procedures were pre-registered, and the full datasets and analysis scripts are available on the Open Science Framework web pages (https://osf.io/24b93/).

Explicit Measures

Gender stereotype endorsement and knowledge. A mixed ANCOVA was performed with participants' Gender Stereotype scores (2: Stereotype Endorsement vs. Stereotype Knowledge) as a within-subjects variable and Participant Gender and Participant Age included as between-subjects factors, whilst controlling for presentation order of the measures. There was no effect of presentation order, $F(1, 59) = 3.36$, $p = .072$, $\eta_p^2 = .05$. In line with hypotheses, there was a significant Gender Stereotype x Participant Age interaction, $F(1, 59) = 21.25$, $p < .001$, $\eta_p^2 = .27$. Post-hoc comparisons (with Bonferroni correction) revealed that, in line with Experiment 1, parents achieved significantly higher scores (M = 7.53, SD = 0.57) than children (M = 3.11, SD = 3.03, p <.001) on the stereotype endorsement measure, indicating greater flexibility around who *should* play with masculine- and feminine-typed toys. However, as expected, the effect of Participant Age on stereotype knowledge scores was non-significant, $p = .143$, indicating that children (M = 2.85, SD = 2.63) and parents (M = 3.98, SD = 2.29) are similarly aware of the gender stereotypes of toys, but children explicitly endorse these stereotypes to a greater extent than parents (i.e. indicate that masculine-typed toys 'should' be played with by boys, and feminine-typed toys 'should' be played with by girls). The pairwise comparison between children's stereotype endorsement and knowledge scores was non-significant; $p = .513$, but this was statistically significant amongst parents; $p < .001$.

As in Experiment 1, the between-subjects effect of Participant Gender and the interaction between Participant Age and Participant Gender were non-significant ($p = .070$; $p = .234$, respectively).

Gender stereotype knowledge and endorsement scores were positively correlated in children, $r = .65$, $p < .001$, but this relationship was non-significant in parents; $r = .16$, $p = .386$.

Implicit Measure – Eye-Tracking

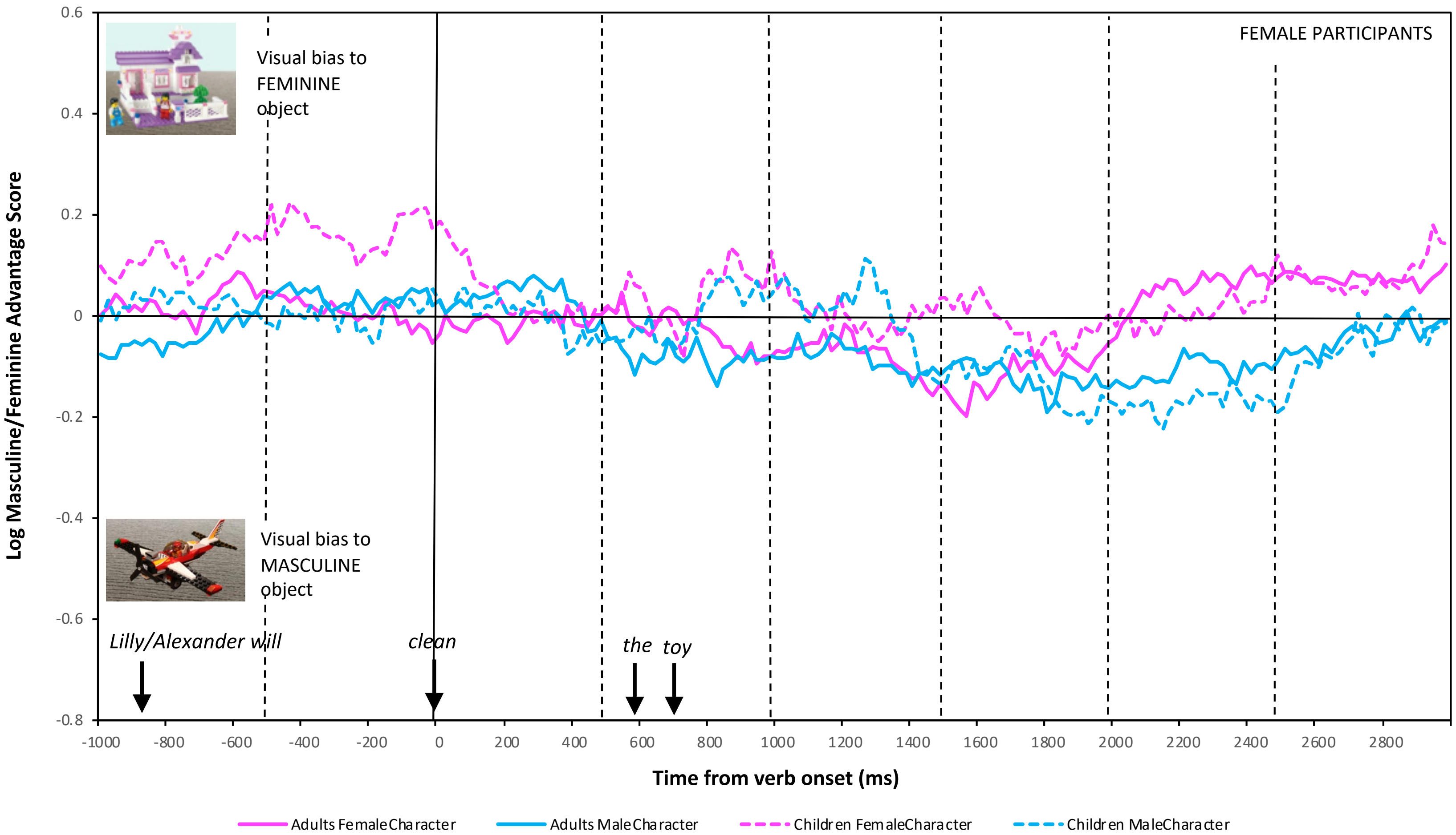
Eye-tracking data was processed as described in Experiment 1. The average masculine/feminine advantage scores in each condition are plotted in Figures 5 and 6, and statistical effects are reported in Table 2.

Most importantly, the effect of Character Gender did not reach significance in any of the time windows around the neutral verb. However, we note that a marginal effect of Character Gender emerged in time window 7 (2500 to 3000ms; $p = .054$), revealing a trend for participants to preferentially fixate the feminine-typed objects in the female-character condition, and the masculine-typed objects in the male-character condition in that time window. In addition, the effect of Participant Gender was significant in time window 8 (2500 to 3000ms), reflecting a looking-preference for the feminine-typed objects among female participants, and a preference for the masculine-typed objects among male participants.

Therefore, overall, participants did not show a preference to fixate the feminine-typed objects in the female-character condition nor the masculine-typed objects in the male-character condition. This is in contrast to Experiment 1 and suggests that when a neutral (rather than an intentional action) verb is included in the audio sentence, participants do not refer to gender stereotyped knowledge of male/female preferences to direct their visual search of objects in the scene. This reinforces the interpretation of gender stereotype effects seen in Experiment 1, suggesting that they are not driven by simple bottom-up processing from informed knowledge about gender associations, but rather they are based on complex gender stereotypes regarding intentional play preferences and behavior.

Figure 5

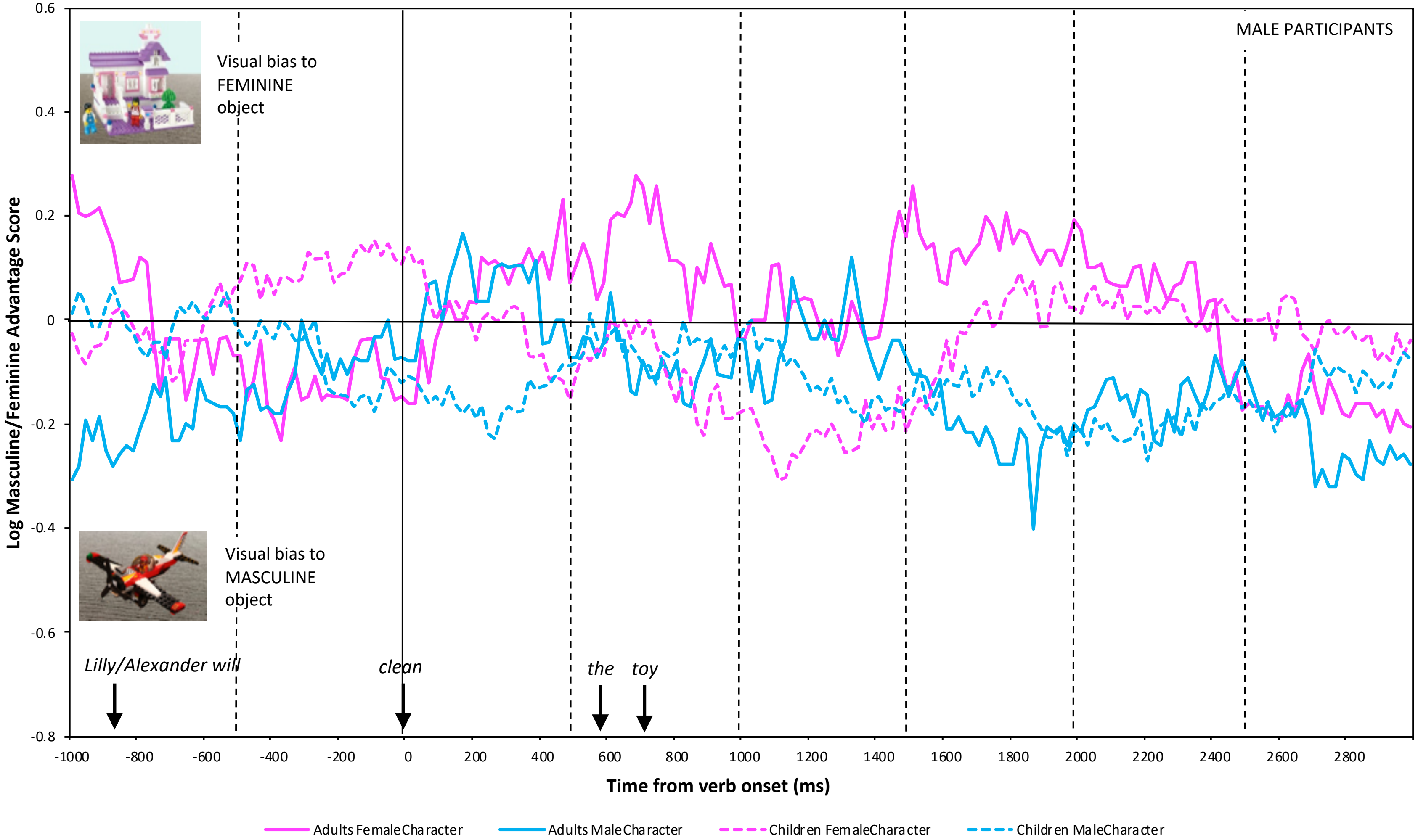
Average location advantage scores for female participants in each condition in Experiment 2.



Note. The solid black vertical line (t=0) indicates the onset of the verb in the audio sentence. The dashed vertical lines represent the 500ms timeslots that were used for statistical analysis.

Figure 6

Average location advantage scores for male participants in each condition in Experiment 2.



Note. The solid black vertical line (t=0) indicates the onset of the verb in the audio sentence. The dashed vertical lines represent the 500ms timeslots that were used for statistical analysis.

Table 2

*Results of Linear Mixed Effects Models for each time window of interest in Experiment 2, where † p < .06, ** p < .05*

		Character Gender	Participant Age	Participant Gender	Character Gender x Participant Age	Character Gender x Participant Gender	Participant Age x Participant Gender	Character Gender x Participant Age x Participant Gender
	Est.	-0.38	0.39	-0.35	0.50	-0.07	0.20	1.14
-1000 to -500ms	*SE*	0.2	0.29	0.24	0.4	0.4	0.44	0.81
	t	-1.88	1.38	-1.42	1.25	-0.17	0.46	1.41
	Est.	-0.27	0.43	-0.50	-0.74	-0.08	0.50	-0.25
-500 to 0ms	*SE*	0.25	0.31	0.28	0.49	0.54	0.51	1.00
	t	-1.09	1.36	-1.79	-1.52	-0.14	0.97	-0.25
	Est	-0.16	-0.28	-0.02	-0.39	-0.30	-0.63	0.10
0 to 500ms	*SE*	0.27	0.26	0.22	0.53	0.52	0.42	1.03
	t	-0.61	-1.09	-0.07	-0.75	-0.58	-1.49	0.09
	Est.	-0.23	0.00	0.01	0.61	-0.09	-0.61	1.42
500 to 1000ms	*SE*	0.23	0.31	0.26	0.47	0.47	0.47	0.93
	t	-1.00	-0.01	0.04	1.32	-0.19	-1.30	1.53
	Est.	0.07	-0.07	-0.25	0.27	0.17	-0.84	0.72
1000 to 1500ms	*SE*	0.28	0.3	0.25	0.54	0.51	0.54	1.22
	t	0.23	-0.25	-1.01	0.51	0.33	-1.55	0.59
	Est.	-0.47	-0.04	0.18	0.20	-0.70	-0.52	1.24
1500 to 2000ms	*SE*	0.29	0.39	0.28	0.54	0.49	0.75	1.17
	t	-1.59	-0.11	0.65	0.37	-1.42	-0.69	1.06
	Est.	-0.71	-0.07	-0.12	0.17	-0.28	0.26	0.42
2000 to 2500ms	*SE*	0.34	0.31	0.28	0.57	0.57	0.67	1.13
	t	-2.09†	-0.21	-0.43	0.30	-0.49	0.38	0.37
	Est.	-0.51	0.34	-0.61	-0.15	0.01	0.64	-0.03
2500 to 3000ms	*SE*	0.32	0.32	0.29	0.50	0.53	0.58	1.13
	t	-1.62	1.07	-2.15**	-0.29	0.03	1.11	-0.03

Note: Data reported here are uncorrected for multiple comparisons. See Experiment 1 "Main analyses" section for rationale.

General Discussion

In this paper we sought to examine whether gender stereotypes about children's toys, activities, and objects are evident in explicit and/or implicit measures, and specifically how they compare between parents and their children. The present experiments are the first to use the visual world paradigm to examine gender stereotypes, building on the limited literature in this area. Our results revealed a dissociation between explicit and implicit measures in parents, where they showed greater gender flexibility than children on the explicit toy-rating measure, but comparably strong implicit gender stereotypes on the eye-tracking measure in Experiment 1. That is, when explicitly questioned about who *should* play with masculine- and feminine-typed toys, parents were more likely to say "both boys and girls" than children, who adhered to gender stereotype norms. This is in line with previous literature which finds that parents are reluctant to reveal gender-stereotypic beliefs as they are socially unacceptable in most western cultures (Axinn et al., 2011). In contrast, when eye-tracking measured implicit gender biases, Experiment 1 showed that both parents and their children exhibited gender biased looking behaviors (i.e. a preference to fixate the feminine objects in the female-character condition, and a preference to fixate the masculine objects in the male-character condition). Crucially, this bias emerged incrementally as the sentence progressed; it was evident immediately following the verb that describes an intentional interaction with the object (e.g. *play with*). This shows that participants made rapid inferences about the intended object, and biased their domain of reference according to stereotypes based on the character's gender.

Importantly, Experiment 2 allowed us to eliminate a low-level explanation for this implicit gender stereotype effect based on gender category activation and semantic associations, as eye movements did not significantly distinguish masculine- and feminine-typed objects when the action verb was replaced with a neutral verb (e.g. *move*). That is, in

Experiment 1, eye movements were prompted by participants' complex gender stereotypes concerning gendered play behavior and intentions of boys and girls. The difference in parents' and children's explicit gender flexibility scores in Experiment 1 was replicated in Experiment 2, whilst differences in gender-stereotype knowledge scores were non-significant between the groups, suggesting that parents and children are similarly aware of gender stereotypes, and that gender stereotype knowledge and endorsement may be two distinct concepts, at least among adults (see Banse et al., 2010; Trautner, Ruble, Cyphers, Kirsten, Behrendt, & Hartmann, 2005). Analyses revealed a positive correlation between stereotype endorsement and stereotype knowledge only in children. The divergence between stereotype endorsement and knowledge in adults is likely due to the development of adults' egalitarian attitudes which are likely to inhibit stereotype endorsement (Liben & Bigler, 2002). In empirical studies and meta-analyses, children's gender-stereotype knowledge has been shown to develop linearly, increasing with age, whereas their endorsement of gender stereotypes is curvilinear; increasing until around the age of 6 years, then decreasing (e.g. Serbin & Sprafkin, 1986; Signorella et al., 1993; Trautner et al., 2005). The children in the present study were 7 years old, thus their developmental age is likely to explain the similarity between their gender stereotype knowledge and endorsement scores. We would expect these to diverge as they got older.

In addition to the strong biases relating to the character's gender in the implicit measure, which influenced object preference throughout the audiovisual scenes, other factors modulated biases across the time windows. Notably, parents showed stronger effects of character gender compared to children at several points throughout the scenes, indicating that implicit gender stereotypes may be more robust in this group. The effect of character gender was also sustained for longer among parents than children, as parents continued to demonstrate visual biases towards gender stereotypical objects several seconds after audio input ceased, whereas this bias declined more rapidly in children. The finding that gender stereotype influences had a stronger and more sustained effect on eye movements among adults than children is important and novel. This contrasts with research on race bias using the IAT, where implicit bias is found to be stable across 6- and 10-year olds, and adults (Baron & Banaji, 2006). The stronger and more sustained implicit gender bias in adults observed in the present research could be due to repeated experience of gender stereotypes in the environment, which could in turn strengthen implicit gender stereotypes over time. An alternative explanation is that using this paradigm, children are more easily distracted from the task and break off fixations on objects earlier, meaning the bias detected is less sustained. Future research should investigate these explanations.

Fazio, Jackson, Dunton, and Williams (1995) and McConnell and Leibold (2001) explain that although explicit attitudes are more closely associated with deliberative behaviors, implicit attitudes are associated with more subtle non-verbal, spontaneous behaviors. The present findings could therefore suggest that even if parents explicitly promote mixed toy play; their strong implicit gender stereotypes could be leaking into non-verbal behavior directed towards their children's toy and activity choices. This could go some way to explain why in studies such as Freeman's (2007), children predict that their parents will possess strong gender stereotypes in relation to toys, despite parents' expressing that they do not endorse gender-typed play: children may be inferring this information from parents' non-verbal behaviors, linked to their implicit gender stereotypes.

In fact, implicit stereotypes have been shown to be better predictors of behavior than explicit stereotypes captured via self-report measures (Greenwald et al., 2009; Kunda & Spencer, 2003; White & White, 2006) which can be affected by social desirability and poor validity due to participants not always being aware of their own stereotypes. Among children, Cvencek et al. (2011a) found that implicit attitudes towards males and females explained more variance in their play preferences than their explicit attitudes did, i.e. the more children implicitly preferred one gender over the other (e.g. female), the greater preference they showed for activities associated with that gender (e.g. feminine-typed toys).

Implicit gender stereotypes observed in relation to the objects in Experiment 1 may translate to other gender-related objects, activities, and behaviors, which could limit girls' and boys' motivation to engage with gender-typed pastimes, school subjects, and careers. It has been shown that strong implicit gender stereotypes in relation to math predict girls' poor performance, leading them to increasingly avoid engagement with the subject (Steffens et al., 2010), and are also associated with larger gaps in actual math performance amongst boys and girls (Nosek et al., 2009). Children's math-gender stereotypes and gender identity have also been found to predict math self-concepts in Singaporean children, with boys identifying more strongly with math than girls, despite girls excelling in Math in this country (Cvencek et al., 2014). Interestingly, Cvencek et al. (2014) employed both implicit and explicit measures and found a strong correlation between scores on these measures in relation to gender identity, but low-moderate correlations in relation to math self-concept and math-gender stereotypes.

The detrimental effect of implicit gender stereotypes on children's performance warrants further investigation; understanding the developmental trajectory of these implicit biases is crucial in order to challenge their formation during early childhood. Future research should examine a wider age range of children to observe how implicit biases develop with age, as well as examining non-parent adults, as explicit measures have shown that this group gender stereotype toys to a greater extent than parents (Campenni, 1999).

The eye-tracking data in Experiment 1 (and to a lesser extent, Experiment 2) also revealed differences in looking preferences according to the participant's own gender. First, the effect of character gender was significantly weaker (or absent) among the male children compared to any of the other groups. In contrast, previous literature has shown that boys have stronger explicit preferences and stereotypes for gender-typed toys than girls do, due to stricter norms around what they can play with (Carter & Levy, 1988; Fagot & Hagan, 1991). Additionally, both parents and children showed persistent gaze biases to own-gender-typed objects. Interestingly, the size of this own-gender bias did not differ between male and female participants. This finding contrasts with previous research in children (e.g. Dunham et al., 2016) and parents (Endendijk et al., 2013) that has demonstrated stronger implicit gender stereotype effects in females than males. Taken together, the findings from Experiment 1 that male children compared to female children showed reduced gender-stereotype biases for others, but comparable own-gender biases, may highlight that children's *stereotypes* and *preferences* are independent constructs (also see Master et al., 2017).

Special attention should also be paid to *male* child and adult implicit stereotypes, as Experiment 1 unveiled significantly stronger biases among male adults than male children, not evident among female participants. Therefore, it may be that boys' implicit gender stereotypes strengthen with age or perhaps during parenthood, but further investigation is required to explore this. It has been documented in early research that fathers impose more rigid sex role expectations on their sons than on their daughters, and that boys' fathers are less flexible than boys' mothers in their definitions of gender appropriate behaviors (Burge, 1981; but see Endendijk, Groeneveld, Bakermans-Kranenburg, & Mesman, 2016). Halpern and Perry-Jenkins (2016) additionally found that mothers and fathers played differential roles in their sons' and daughters' acquisition of gender-stereotype knowledge. Although there is some research which examines the relationships between explicit and implicit gender stereotypes among mothers/fathers and sons/daughters, e.g. del Río et al. (2018; in relation to math), future research should continue to explore how implicit gender stereotypes are influenced by familial relationships, particularly amongst male adults and children, and in other gender-stereotype domains, e.g. toy-play. It is worthwhile noting the possibility that boys in the present study may have engaged with the task to a lesser extent than girls, which could explain weaker biases due to inattention.

Limitations

As with most studies in the field, some caution must be applied when attempting to generalize findings. The sample was homogeneous, predominately recruited from a White, middle-upper income area. This is important to consider when investigating gender stereotypes as parents' socioeconomic status can be an influencing factor: traditional gender roles are more likely to be endorsed by those from lower-income backgrounds (see Mesman & Groeneveld, 2018, for discussion). Future research would benefit from collecting additional demographic data on education levels and socioeconomic status, and recruiting a more diverse sample to better reflect various cross-sections of society, including those from non-Western backgrounds, and fathers, who are typically under-represented in this research area.

There is also the possibility that the similarity of children's gender stereotype knowledge and endorsement scores was due to a lack of understanding of the differences between the phrasing of the questions. Ceiling effects were seen in the parents' gender flexibility scores, which may be due to the limited number of items (eight) contained in the gender stereotype endorsement measure. Future research would benefit from using a larger number of items and/or a more incremental scale to allow for a more nuanced exploration of explicit stereotypes, and possibly greater variation in participants' responses. The ability to examine relationships between implicit and explicit data would also be beneficial.

Additionally, the objects included in the audiovisual scenes were chosen because they clearly adhered to masculine and feminine stereotypes in terms of both form and color, to ensure they were realistic and easily recognizable to children and parents. However, as a result we cannot differentiate between gender stereotypes associated with objects and colors as they are confounded. It would be of interest for future research to experimentally manipulate these features to examine them in isolation.

Finally, the experiments presented here focus on the role of parents in children's gender stereotype development, but it is important to also acknowledge the significant influence of peers in this domain, particularly from middle-childhood onwards (Martinez, Osornio, Halim, & Zosuls, 2020). Whilst some parents may explicitly discourage gender stereotyping, it is well documented that peers encourage adherence to gender norms via same-sex play and policing of behavior and appearance, and are therefore key players in reinforcing stereotypes (Martinez et al., 2020). It is important for future research to examine the contributing role of socializing agents such as peers, as well as parents, on children's implicit and explicit gender stereotypes.

Future Research

Future research should also attempt to re-train implicit associations in children. Research that has attempted this with adults has shown that re-training math-gender implicit associations can have a buffering effect during stereotype threat conditions. Women who were re-trained to associate 'liking' with mathematics demonstrated more effort and higher working memory during a math task, especially when gender stereotypes were salient (Forbes & Schmader, 2010). A recent study by Master et al. (2017) tested an intervention among 6-year-old children in which they learned to program robots. This experience led to increased explicit STEM interest and self-efficacy among girls (in comparison to a control condition), and closed the gender gap with boys on these measures. It would be interesting to examine whether interventions such as Master's et al. (2017) also lead to a change in children's implicit gender stereotypes. It is logical that interventions to re-train implicit associations may be even more successful with young children than adults, as gender stereotypes would be less culturally ingrained (Baron, Schmader, Cvencek, & Meltzoff, 2014).

Conclusions

In conclusion, we have presented two experiments (one pre-registered) that make a significant contribution to our understanding of gender stereotypes in parents and their children, and introduce a well-established paradigm to this field for the first time. Most importantly, we identified a discrepancy between parents' implicit and explicit gender stereotypes, which was not present in their children. Parents showed greater explicit gender-toy flexibility than children, but strong implicit gender biases on the eye-tracking measure. In fact, these implicit gender biases were stronger and more sustained among the parents than children. This suggests that the implicit measure used in these experiments, the visual world paradigm, is able to successfully capture peoples' unconscious gender biases which may be masked on explicit measures. The incongruity between parents' implicit and explicit gender stereotypes might suggest that these constructs begin to diverge with age. Further research is needed to unravel the complex relation between implicit and explicit gender stereotypes and their influences on attitudes, behavior, and preferences, particularly from a developmental perspective.

Acknowledgements

The authors sincerely thank the research assistants in the School of Psychology at the University of Kent who assisted with participant recruitment and data collection. This work was supported by The University of Kent's 50th Anniversary PhD Scholarship Award and the School of Psychology's Seed Fund.

References

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247-264.

Altmann, G. T., & Kamide, Y. (2007). The real-time mediation of visual attention by language and world knowledge: Linking anticipatory (and other) eye movements to linguistic processing. *Journal of Memory and Language, 57*(4), 502-518. doi:10.1016/j.jml.2006.12.004

Altmann, G. T., & Kamide, Y. (2009). Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition, 111*(1), 55-71. doi:10.1016/j.jml.2006.12.004

Axinn, W. G., Young-DeMarco, L., & Ro, M. C. (2011). Gender double standards in parenting attitudes. *Social Science Research*, *40*(2), 417-432. doi: 10.1016/j.ssresearch.2010.08.010

Bandura, A., & Bussey, K. (2004). On broadening the cognitive, motivational, and sociostructural scope of theorizing about gender development and functioning: Comment on Martin, Ruble, and Szkrybalo (2002). *Psychological Bulletin, 130*(5), 691-701. doi:10.1037/0033-2909.130.5.691

Banse, R., Gawronski, B., Rebetez, C., Gutt, H., & Bruce Morton, J. (2010). The development of spontaneous gender stereotyping in childhood: Relations to stereotype knowledge and stereotype flexibility. *Developmental Science*, *13*(2), 298-306. doi:10.1111/j.1467-7687.2009.00880.x

Baron, A. S., & Banaji, M. R. (2006). The development of implicit attitudes: Evidence of race evaluations from ages 6 and 10 and adulthood. *Psychological Science, 17*(1), 53-58. doi:10.1111/j.1467-9280.2005.01664.x

Baron, A. S., Schmader, T., Cvencek, D., & Meltzoff, A. N. (2014). The gendered self-concept: How implicit gender stereotypes and attitudes shape self-definition. In P. J. Leman & H. R. Tenenbaum (Eds*.), Current issues in developmental psychology. Gender and development* (pp. 109-132). New York, NY, US: Psychology Press

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255-278. doi:10.1016/j.jml.2012.11.001

Black, J., Barzy, M., Williams, D., & Ferguson, H. (2019). Intact counterfactual emotion processing in autism spectrum disorder: Evidence from eye- tracking. *Autism Research*, *12*(3), 422-444. doi:10.1002/aur.2056

Berends, S. M., Brouwer, S. M., & Sprenger, S. A. (2016). Eye-tracking and the visual world paradigm. In M. S. Schmid, S. M. Berends, C. Bergmann, S. M. Brouwer, N. Meulman, B. J. Seton, S. A. Sprenger, & L. A. Stowe, *Designing research on bilingual development* (pp. 55-80). USA: Springer.

Bian, L., Leslie, S. J., & Cimpian, A. (2017). Gender stereotypes about intellectual ability emerge early and influence children's interests. *Science*, *355*(6323), 389-391. doi: 10.1126/science.aah6524

Blakemore, J. O., Berenbaum, S. A., & Liben, L. S. (2008). *Gender Development***.** New York: Psychology Press.

Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles, 53*(9), 619-633. doi: 10.1007/s11199-005-7729-0.

Burge, P. (1981). Parental Child- Rearing Sex- Role attitudes related to social issue Sex-Role attitudes and selected demographic variables. *Family and Consumer Sciences Research Journal, 9*(3), 193-199.

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review, 106*(4), 676-713. doi:10.1037/0033-295X.106.4.676

Callan, M. J., Ferguson, H. J., & Bindemann, M. (2013). Eye movements to audiovisual scenes reveal expectations of a just world. *Journal of Experimental Psychology: General, 142*(1), 34. doi:10.1037/a0028261

Campenni, C. E. (1999). Gender stereotyping of children's toys: A comparison of parents and nonparents. *Sex Roles, 40*(1-2), 121-138. doi:10.1023/A:1018886518834

Carter, D. B., & Levy, G. D. (1988). Cognitive aspects of early sex-role development: The influence of gender schemas on preschoolers' memories and preferences for sex-typed toys and activities. *Child Development,* 782-792. doi:10.2307/1130576

Cherney, I. D., Kelly-Vance, L., Glover, K. G., Ruane, A., & Ryalls, B. O. (2003). The effects of stereotyped toys and gender on play assessment in children aged 18-47 months. *Educational Psychology, 23*(1), 95-105. doi:10.1080/01443410303222

Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles, 54*(9-10), 717-726. doi:10.1007/s11199-006-9037-8

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*(1), 84-107. doi:10.1016/0010-0285(74)90005-X

Coyne, S. M., Linder, J. R., Rasmussen, E. E., Nelson, D. A., & Collier, K. M. (2014). It’s a bird! It’s a plane! It’s a gender stereotype!: Longitudinal associations between superhero viewing and gender stereotyped play. *Sex Roles*, *70*(9-10), 416-430. doi:10.1007/s11199-014-0374-8

Coyne, S. M., Linder, J. R., Rasmussen, E. E., Nelson, D. A., & Birkbeck, V. (2016). Pretty as a princess: Longitudinal effects of engagement with disney princesses on gender stereotypes, body esteem, and prosocial behavior in children. *Child Development, 87*(6), 1909-1925. doi:10.1111/cdev.12569

Crosby, J. R., Monin, B., & Richardson, D. (2008). Where do we look during potentially offensive behavior? *Psychological Science, 19*(3), 226-228. doi:10.1111/j.1467-9280.2008.02072.x

Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011b). Math–gender stereotypes in elementary school children. *Child Development*, *82*(3), 766-779. doi:10.1111/j.1467-8624.2010.01529.x.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2011a). Measuring implicit attitudes of 4-year-olds: The preschool implicit association test. *Journal of Experimental Child Psychology, 109*(2), 187-200. doi:10.1016/j.jecp.2010.11.002

Cvencek, D., Meltzoff, A. N., & Kapur, M. (2014). Cognitive consistency and math–gender stereotypes in Singaporean children. *Journal of experimental child psychology*, *117*, 73-91. doi:10.1016/j.jecp.2013.07.018

Cvencek, D., Kapur, M., & Meltzoff, A. N. (2015). Math achievement, stereotypes, and math self-concepts among elementary-school students in Singapore. *Learning and Instruction*, *39*, 1-10. doi:10.1016/j.learninstruc.2015.04.002

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology: Research and Theory on Human Development, 163*(3), 272-282. doi:10.1080/00221320209598683

del Río, M. F., Strasser, K., Cvencek, D., Susperreguy, M. I., & Meltzoff, A. N. (2018). Chilean kindergarten children’s beliefs about mathematics: Family matters. *Developmental Psychology* (in press). doi:10.1037/dev0000658

Dunham, Y., Baron, A. S., & Banaji, M. R. (2016). The development of implicit gender attitudes. *Developmental Science, 19*(5), 781-789. doi:10.1111/desc.12321

Else-Quest, N. M., Hyde, J. S., Goldsmith, H. H., & Van Hulle, C. A. (2006). Gender differences in temperament: a meta-analysis. *Psychological Bulletin*, *132*(1), 33. doi: 10.1037/0033-2909.132.1.33

Endendijk, J. J., Groeneveld, M. G., Bakermans-Kranenburg, M. J., & Mesman, J. (2016). Gender-differentiated parenting revisited: Meta-analysis reveals very few differences in parental control of boys and girls. *PloS One, 11*(7). doi:10.1371/journal.pone.0159193

Endendijk, J. J., Groeneveld, M. G., van Berkel, S. R., Hallers-Haalboom, E. T., Mesman, J., & Bakermans-Kranenburg, M. J. (2013). Gender stereotypes in the family context: Mothers, fathers, and siblings. *Sex Roles, 68*(9-10), 577-590. doi:10.1007/s11199-013-0265-4

Fagot, B. I., & Hagan, R. (1991). Observations of parent reactions to sex- stereotyped behaviors: Age and sex effects. *Child Development, 62*(3), 617-628. doi:10.2307/1131135

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69*(6), 1013. doi:10.1037/0022-3514.69.6.1013

Ferguson, H. J., & Breheny, R. (2011). Eye movements reveal the time-course of anticipating behavior based on complex, conflicting desires. *Cognition, 119*(2), 179-196. doi:10.1016/j.cognition.2011.01.005

Ferguson, H. J., & Breheny, R. (2012). Listeners' eyes reveal spontaneous sensitivity to others' perspectives. *Journal of Experimental Social Psychology, 48*(1), 257-263. doi:10.1016/j.jesp.2011.08.007

Ferguson, H. J., Scheepers, C., & Sanford, A. J. (2010). Expectations in counterfactual and theory of mind reasoning. *Language and Cognitive Processes, 25*(3), 297-346. doi:10.1016/j.jml.2006.12.004

Fitzroy, S., & Rutland, A. (2010). Learning to control ethnic intergroup bias in childhood. *European Journal of Social Psychology, 40*(4), 679-693. doi:10.1002/ejsp.746

Forbes, C. E., & Schmader, T. (2010). Retraining attitudes and stereotypes to affect motivation and cognitive capacity under stereotype threat. *Journal of Personality and Social Psychology, 99*(5), 740. doi:10.1037/a0020971

Freeman, N. K. (2007). Preschoolers' perceptions of gender appropriate toys and their parents' beliefs about genderized behaviors: Miscommunication, mixed messages, or hidden truths? *Early Childhood Education Journal, 34*(5), 357-366. doi:10.1007/s10643-006-0123-x

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, *7*(4), 493-498. doi:10.1111/2041-210X.12504

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74*(6), 1464. doi:10.1037//0022-3514.74.6.1464

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-Analysis of Predictive Validity. *Journal of Personality and Social Psychology, 97*(1), 17-41. doi: 10.1037/a0015575

Halpern, H. P., & Perry-Jenkins, M. (2016). Parents' gender ideology and gendered behavior as predictors of children's gender-role attitudes: A longitudinal exploration. *Sex Roles, 74*(11-12), 527-542. doi:10.1007/s11199-015-0539-0

Hastings, P. D., McShane, K. E., Parker, R., & Ladha, F. (2007). Ready to make nice: Parental socialization of young sons' and daughters' prosocial behaviors with peers. *The Journal of Genetic Psychology*, *168*(2), 177-200. doi:10.3200/GNTP.168.2.177-200

Hewstone, M., Rubin, M., & Willis, H. (2002). Intergroup bias. *Annual Review of Psychology, 53*(1), 575-604. doi:10.1146/annurev.psych.53.100901.135109

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the implicit association test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31*(10), 1369-1385. doi:10.1177/0146167205275613

Idle, T., Wood, E., & Desmarais, S. (1993). Gender role socialization in toy play situations: Mothers and fathers with their sons and daughters. *Sex Roles, 28*(11-12), 679-691. doi:10.1007/BF00289987

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26*(3), 302-310. doi:10.1177/0956797614563338

Kamide, Y., Altmann, G. T., & Haywood, S. L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and language*, *49*(1), 133-156. doi: 10.1016/S0749-596X(03)00023-8

Kamide, Y., Lindsay, S., Scheepers, C., & Kukona, A. (2016). Event processing in the visual world: Projected motion paths during spoken sentence comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(5), 804. doi:10.1037/xlm0000199

Kollmayer, M., Schultes, M. T., Schober, B., Hodosi, T., & Spiel, C. (2018). Parents' judgments about the desirability of toys for their children: Associations with gender role attitudes, gender-typing of toys, and demographics. *Sex Roles, 79*(5-6), 329-341. doi:10.1007/s11199-017-0882-4

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin, 129*(4), 522. doi:10.1037/0033-2909.129.4.522

Li, R. Y. H., & Wong, W. I. (2016). Gender-typed play and social abilities in boys and girls: Are they related? *Sex Roles, 74*(9-10), 399-410. doi:10.1007/s11199-016-0580-7

Liben, L. S., & Bigler, R. S. (2002). *The development course of gender differentiation*. Blackwell publishing.

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. doi:10.1111/j.0963-7214.2004.00276.x

Martin, C. L., Ruble, D. N., & Szkrybalo, J. (2002). Cognitive theories of early gender development. *Psychological bulletin, 128*(6), 903-933. doi: 10.1037//0033-2909.128.6.903

Martinez, M.A., Osornio, A., Halim, M.L., & Zosuls, K.M. (2020). Gender: Awareness, identity, and stereotyping. In J.B. Benson (Ed.), *Encyclopedia of infant and early childhood development* (pp. 1-12). Elsevier.

Master, A., Cheryan, S., Moscatelli, A., & Meltzoff, A. N. (2017). Programming experience promotes higher STEM motivation among first-grade girls. *Journal of Experimental Child Psychology*, *160*, 92-106. doi:10.1016/j.jecp.2017.03.013

McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology, 37*(5), 435-442. doi:10.1006/jesp.2000.1470

McHale, S. M., Crouter, A. C., & Tucker, C. J. (1999). Family context and gender role socialization in middle childhood: Comparing girls to boys and sisters to brothers. *Child Development, 70*(4), 990-1004. doi:10.1111/1467-8624.00072

McHale, S. M., Crouter, A. C., & Whiteman, S. D. (2003). The family contexts of gender development in childhood and adolescence. *Social Development, 12*(1), 125-148. doi:10.1111/1467-9507.00225

Mesman, J., & Groeneveld, M. G. (2018). Gendered parenting in early childhood: Subtle but unmistakable if you know where to look. *Child Development Perspectives*, *12*(1), 22-27. doi:10.1111/cdep.12250

Meyer, M., & Gelman, S. A. (2016). Gender essentialism in children and parents: Implications for the development of gender stereotyping and gender-typed preferences. *Sex Roles, 75*(9-10), 409-421. doi:10.1007/s11199-016-0646-6

Mirković, J., & Altmann, G. T. (2019). Unfolding meaning in context: The dynamics of conceptual similarity. *Cognition*, *183*, 19-43. doi:10.1016/j.cognition.2018.10.018

Mischel, W. (1966). A social learning view of sex differences in behavior. In V. Burr (Ed.), *Social Constructionism* (pp. 108-129). New York, US: Routledge/Taylor & Francis Group. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2015-16208-007&site=ehost-live

Nock, M. K., & Banaji, M. R. (2007). Prediction of suicide ideation and attempts among adolescents using a brief performance-based test. *Journal of Consulting and Clinical Psychology*, *75*(5), 707-715. doi:10.1037/0022-006X.75.5.707

Nosek, B. A. (2005). Moderators of the relationship between implicit and explicit evaluation. *Journal of Experimental Psychology: General*, *134*(4), 565-584. doi:10.1037/0096-3445.134.4.565

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math= male, me= female, therefore math≠ me. *Journal of Personality and Social Psychology, 83*(1), 44. doi:10.1037/0022-3514.83.1.44

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., ... & Kesebir, S. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences*, *106*(26), 10593-10597. doi: 10.1073/pnas.0809921106

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J., & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of IAT criterion studies. *Journal of Personality and Social Psychology*, *105*(2), 171. doi:10.1037/a0032734

Paoletti, J. (2012). *Pink and blue: Telling the boys from the girls in America*. Bloomington, IN: Indiana University Press. Retrieved from http://search.ebscohost.com/ login.aspx?direct=true&db=psyh&AN=2012-08701-000&site=ehost-live

Pomerleau, A., Bolduc, D., Malcuit, G., & Cossette, L. (1990). Pink or blue: Environmental gender stereotypes in the first two years of life. *Sex Roles, 22*(5-6), 359-367. doi:10.1007/BF00288339

Rutland, A., Cameron, L., Milne, A., & McGeorge, P. (2005). Social norms and self-presentation: Children's implicit and explicit intergroup attitudes. *Child Development, 76*(2), 451-466. doi: 10.1111/j.1467-8624.2005.00856.x

Serbin, L. A., Poulin-Dubois, D., Colburne, K. A., Sen, M. G., & Eichstedt, J. A. (2001). Gender stereotyping in infancy: Visual preferences for and knowledge of gender-stereotyped toys in the second year. *International Journal of Behavioral Development, 25*(1), 7-15. doi: 10.1080/01650250042000078

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

Sherman, A. M., & Zurbriggen, E. L. (2014). “Boys can be anything”: Effect of Barbie play on girls’ career cognitions. *Sex Roles, 70*(5-6), 195-208. doi: 10.1007/s11199-014-0347-y

Signorella, M. L., Bigler, R. S., & Liben, L. S. (1993). Developmental differences in children′s gender schemata about others: A meta-analytic review. *Developmental review*, *13*(2), 147-183. doi:10.1006/drev.1993.1007

Spinner, L., Cameron, L., & Calogero, R. (2018). Peer toy play as a gateway to children’s gender flexibility: The effect of (counter)stereotypic portrayals of peers in children’s magazines. *Sex Roles, 79*(5-6), 314-328. doi:10.1007/s11199-017-0883-3

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: Comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology, 102*(4), 947-963. doi:10.1037/a0019920

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science,* 1632-1634.

Tenenbaum, H. R., & Leaper, C. (2002). Are parents' gender schemas related to their children's gender-related cognitions? A meta-analysis. *Developmental Psychology, 38*(4), 615-630. doi:10.1037/0012-1649.38.4.615

Todd, B. K., Fischer, R. A., Di Costa, S., Roestorf, A., Harbour, K., Hardiman, P. & Barry, J. A. (2017). Sex differences in children’s toy preferences: A systematic review, meta-regression, and meta-analysis. *Infant & Child Development*, 27, 1-29. doi: 10.1002/icd.2064

Trautner, H. M., Ruble, D. N., Cyphers, L., Kirsten, B., Behrendt, R., & Hartmann, P. (2005). Rigidity and flexibility of gender stereotypes in childhood: Developmental or differential?. *Infant and Child Development, 14*(4), 365-381. doi:10.1002/icd.399

Weisgram, E. S., Fulcher, M., & Dinella, L. M. (2014). Pink gives girls permission: Exploring the roles of explicit gender labels and gender-typed colors on preschool children's toy preferences. *Journal of Applied Developmental Psychology, 35*(5), 401-409. doi:10.1016/j.appdev.2014.06.004

White, M. J., & White, G. B. (2006). Implicit and explicit occupational gender stereotypes. *Sex Roles, 55*(3-4), 259-266. doi:10.1007/s11199-006-9078-z

Wong, W. I., & Hines, M. (2015). Preferences for pink and blue: The development of color preferences as a distinct gender-typed behavior in toddlers. *Archives of Sexual Behavior, 44*(5), 1243-1254. doi:10.1007/s10508-015-0489-1

Wood, E., Desmarais, S., & Gugula, S. (2002). The impact of parenting experience on gender stereotyped toy play of children. *Sex Roles, 47*(1-2), 39-49. doi:10.1023/A:1020679619728

Appendix A: Experimental Scenes and Audio Sentences

"Sarah/Ben has drawn/seen a picture."

"Emily/Thomas will arrange/use the stickers."

"Evie/Jack has dressed/will trip over the toy."

"Chloe/Oliver will wear/wash a costume."

"Lilly/Alexander will play with/clean the toy."

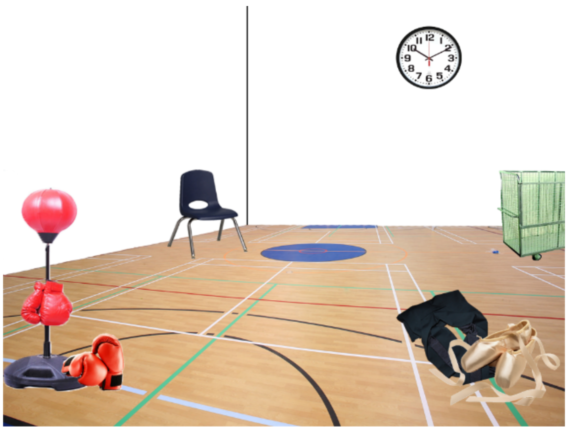
"Sophie/Harry has watched/dropped a film."

"Megan/Edward has built/tripped over a Lego model."

"Hannah/Alfie will wear/tidy away the shoes."

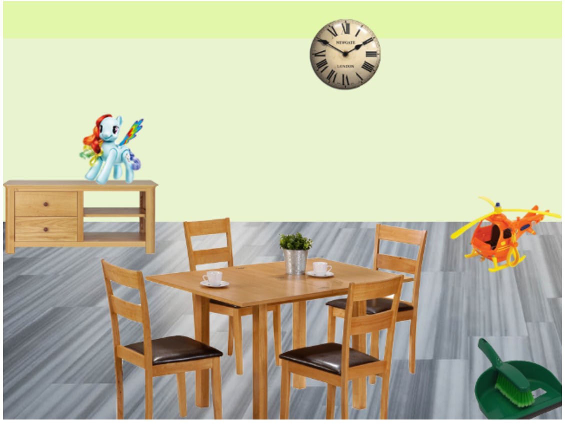
"Lucy/Charlie will play with/throw the toy."

"Emma/David has baked some/noticed the cookies."



"Mollie/Louis will practice/hear about a hobby."

"Rosie/Daniel will ride/kick the bike."



“Millie/William will play with/knock over the toy.”

“Jenny/Ethan has made/found a bracelet.”

“Gemma/Joshua will read/stand on a book.”

“Rachel/Matthew will wear/move the hat.”

Appendix B: Filler Scenes and Audio Sentences

“The food will be eaten by the dog.”

“The cat will climb a tree.”

“The bird has made a nest.”

“The postman has delivered a parcel.”

“The ball will be kicked by the rugby player.”

“The machine will be driven by the farmer.”

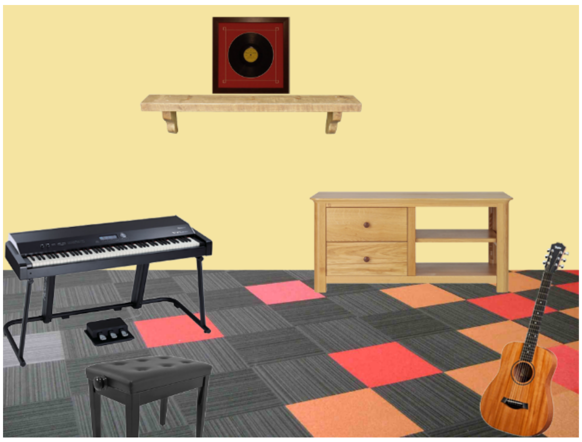
"Rachel will paint with a brush."

"Dr Davis will put on the jacket."

"The board will be wiped by Mrs. Jones."

"The vegetables will be peeled by Michael."

"The plants will be watered by Paul."



"Laura will play an instrument."

"The tool was carried by Johnny."

"The swimmer will wear the goggles."

"The hat will be worn by the builder."

"The woman will drink the wine."