Boys’ toys, girls’ toys: An fMRI study of mothers’ neural responses to children violating gender expectations

Joyce J. Endendijk, Anne K. Smit, Anneloes L. van Baar, Peter A. Bos

1. Introduction

Gender is an important organizer of social life (Blakemore, Berenbaum, & Liben, 2009). One of the first questions a pregnant couple has in daily life is, whether the baby will be a boy or a girl, which is known to guide parents’ choices with regard to names (Barry & Harper, 1995), toys, clothes, books, and design of the newborn’s room (Pomerleau, Bolduc, Malcuit, & Cossette, 1990). Gender also plays an important role in face processing, which is evident, for example, in findings that 3- to 4-month-old infants are already able to discriminate between men and women on the basis of internal facial features (Quinn, Yahr, Kuhn, Slater, & Pascalis, 2002). In adults, gender categorization appears to be a fast, relatively automatic process (e.g., around 130 ms; Clutterbuck & Johnston, 2004; Hügelschafer, Jaudas, & Achtziger, 2016), that can lead to the activation of gender stereotypical knowledge (e.g., women like pink), in particular when people encounter unfamiliar faces (Habibi & Khurana, 2012). Such processes could explain why, from the moment children are born, their parents, but also other people in the child’s surroundings, have vastly different expectations about the characteristics, interests, future roles, and behaviors of boys and girls. For example, parents have been found to associate toys like cars, fire trucks, and tools more with boys, whereas they associated dolls, Barbie, and toy kitchens more with girls (Endendijk et al., 2013, 2014). Parents also expect boys to be more athletic and interested in sports than girls (Fredricks & Eccles, 2002).

These stereotyped expectations can have a powerful effect on parents’ perception of and behavior toward children (Wood & Eagly, 2012). For instance, parents with stereotyped expectations about boys and girls have been found to respond more negatively to children’s behavior that violates gender-role expectations (Endendijk et al., 2014; Friedman, Leaper, & Bigler, 2007; Sandnabba & Ahlberg, 1999), and in particular to boys’ gender atypical behavior (Kane, 2006; Sandnabba & Ahlberg, 1999). Such experiences are likely to shape children’s own expectations about gender (Endendijk et al., 2013), as well as their behavior, via self-fulfilling processes in which children are actually going to behave according to these social expectations (Wood & Eagly, 2012). More specifically, parents stereotyped expectations about boys and girls have been associated with gender-inequality, the development of traditional gender stereotypes in children (Turner & Gervai, 1995), and the development of gender differences in child problem behavior (Endendijk, Groeneveld et al., 2015).

In order to prevent the self-fulfilling cycle and intergenerational
transmission of gender stereotypes, we need a better understanding of the mechanisms underlying parents’ responses to child behavior violating gendered expectations. Therefore, in the current study we examined mothers’ neural responses to pictures of boys and girls who conform or violate social expectations regarding toy preferences. We also examined whether neural processing of these gender-stereotype-confirming and -violating stimuli interacts with child gender, and mothers’ gender stereotypes and beliefs.

There are several indications that neural processes might play a role in people’s responses to stereotype-violating persons. For instance, adults as well as infants direct their attention toward males and females in activities that were inconsistent with gender stereotypes (Hutter, Crisp, Humphreys, Waters, & Moffitt, 2009; Serbin, Poulin-Dubois, & Eichstedt, 2002). In addition, stereotype-violating persons elicit executive processes such as inconsistency resolution and reorganization of memory around the specific individual encountered (Macrae, Bodenhausen, Schoorscheidt, & Milne, 1999). Furthermore, previous fMRI research on the neural correlates of violations of social expectations has demonstrated involvement of the following brain regions: precentral, anterior cingulate cortex, dorsolateral and dorsomedial prefrontal cortex (dIPFC, dmPFC), and the tempoparietal junction (TPJ); areas involved in conflict monitoring, behavioral control, and social cognition (Cloutier, Gabrieli, O’young, & Ambady, 2011; Hehman, Ingbretsen, & Freeman, 2014; Li, Cardenas-Iniguez, Correll, & Cloutier, 2016; Quadflieg et al., 2011). However, only one of these studies has examined violations of gender expectations, but with adult targets and in a non-parent sample (Quadflieg et al., 2011). In this study, participants showed increased dlPFC activation when making social judgements of pictures of men and women who violated gender stereotypes about occupations (male florist, female football player) compared to individuals who confirmed gender stereotypes. We do not know whether similar activation patterns can be found for mothers’ neural responses to child stimuli that confirm or violate gender expectations about toys. This is important to study because child stimuli elicit different neural responses than adult stimuli in parents (Leibenluft, Gobhini, Harrison, & Haxby, 2004) as well as non-parents (Caria et al., 2012). We also do not know whether there are differences in the neural responses to boys and girls who violate gender expectations. This could shed light on why parents discourage behavior that violates gender-role expectations in boys more than in girls (Kane, 2006; Sandnabba & Ahlberg, 1999).

Of the above-mentioned areas, the TPJ, dmPFC, dIPFC, and ACC fit with neural models of stereotypes (Amiodo, 2014; Stanley, Phelps, & Banaji, 2008). The TPJ is supposed to be involved in stereotype-related expectations about social groups, because of its role in understanding the goals or intentions behind behaviors of other people (Amiodo, 2014; Van Overwalle, 2009; Saxe, 2006; Young, Dodell-Feder, & Saxe, 2010). Areas in the medial PFC are thought to be involved in stereotyping via the formation of a repository of social knowledge over time (Van Overwalle, 2009) or via the formation of an impression of an individual’s traits, preferences, and mental states based on stereotyped expectations (Cloutier et al., 2011; Li et al., 2016; Rushworth, Buckley, Behrens, Walton, & Bannerman, 2007). The ACC is presumed to monitor conflict between people’s social expectations and incoming social information (Alexander & Brown, 2011; Botvinick, Cohen, & Carter, 2004; Stanley et al., 2008). As the conflict signal in the ACC rises, the ACC increasingly engages the dlPFC to regulate behavior and inhibit stereotype-based responses (Amiodo, 2014; Stanley et al., 2008). In line with this model, activation in the dlPFC, dmPFC, and ACC is found when non-parents are asked to pair traits to men and women in both a stereotype-congruent and stereotype-incongruent way during an Implicit Association Test (IAT; Knutson, Mah, Manly, & Graffman, 2007), or when non-parents are asked whether certain traits are expected more from men, from women, or from both equally (Mitchell, Ames, Jenkins, & Banaji, 2009; Quadflieg et al., 2009). Moreover, transcranial magnetic stimulation over the left dlPFC and right dmPFC has been found to increase gender-stereotyped responding of non-parents during an IAT, suggesting that these areas play a direct role in controlling gender-stereotypical responses (Cattaneo, Mattavelli, Platania, & Papagno, 2011).

It is not likely that all mothers’ process stimuli that violate or confirm gender stereotypes in the same way, because people differ in stereotype expectations and motivation to respond without prejudice. Previous research using event related potentials has demonstrated that brain activity registering gender-based incongruence of toy-child pairings interacted with mothers’ gender stereotypes (Endendijk, Groeneweld et al., 2018, 2018b). In addition, fMRI research on racial stereotypes demonstrated that high self-reported internal motivation to respond without prejudice attenuated the greater dmPFC activity to Black and White individuals who violated racial attitudes, compared to individuals whose behavior confirmed racial stereotypes (Li et al., 2016). Relatedly, adults with strong racial stereotypes about emotions showed stronger dlPFC activation to individuals whose behavior violated racial stereotypes (Hehman et al., 2014). These findings seem to suggest that more egalitarian stereotypes or a strong motivation to behave in a non-prejudiced way may attenuate people’s sensitivity to stereotype-incongruence, thus being associated with a relative diminished differential dmPFC/dlPFC activity to incongruent versus congruent targets (Li et al., 2016; Sherman et al., 2008).

The current study examined mothers’ neural responses to pictures of boys and girls who confirm or violate social expectations regarding toy preferences. First, we expected that exposure to targets violating gender expectations for toy preferences would increase neural activation in the ACC, dlPFC, dmPFC, and TPJ. Second, we expected that these activation patterns would be stronger for boy targets than for girl targets. In addition, this study explored how variability in mothers’ orientation to gender (e.g., gender stereotypes, motivation to act without prejudice) interacts with neural responses to social targets that violate versus confirm social expectations regarding toy preferences during impression formation.

2. Material and methods

2.1. Participants

A total number of 23 right-handed mothers with one or more children, aged 2–6 years old, were recruited via the university website, parenting websites, and leaflets handed out in child-care centers. The age range of the children corresponds with the developmental phase in which parents play an important role in children’s gender development (McHale, Clouter, & Whiteman, 2003). Participants had no history of psychotic, neurological, or endocrine abnormalities. Participants were not pregnant and did not use psychotropic medication. See Table 1 for the demographic characteristics of the mothers and children. Previous research using a similar violations to social expectations paradigm as in the current study, showed that a sample size of 20 was sufficient to detect moderate to large effects (Cloutier et al., 2011; Hehman et al., 2014).

2.2. Procedure

Scanning sessions were scheduled within 5–10 days following the start of menstruation to diminish endocrine variation due to cyclic effects. Participants were informed not to drink alcohol or use drugs 24 h prior to study participation. Before the scan session participants were screened for MRI contraindications, and alcohol and drug use, and they were given brief explanations of the task. All participating mothers gave written informed consent. Next, participants were screened using a metal detector, and they were instructed to position themselves on the scanner bed as comfortable as possible and to try to relax. Head movement was minimized by foam pads, which were placed between the radiofrequency (RF)-coil and participant’s head. Instructions and
2.3. Measures

2.3.1. Impression formation task

2.3.1.1. Stimuli. A stimulus set consisting of 20 Caucasian child faces (10 girl, 10 boy) with neutral facial expressions was used. These stimuli were selected from the CAFE set of 2–8-year-old children (LoBue & Thrasher, 2015). A group of 44 university students rated each picture for gender-typicality of the child’s appearance on a scale from 1 (clear male-typed appearance) to 5 (clear female-typed appearance). We selected 10 pictures of boys and 10 pictures of girls that were rated most clearly as male-typed ($M = 4.62, SD = 0.32$) and female-typed ($M = 1.29, SD = 0.33$) and female-typed ($M = 1.29, SD = 0.33$). Boy pictures were rated significantly more masculine ($t(43) = -34.38, p < .001$) than the neutral midpoint of the scale (3), and girl pictures were rated significantly more feminine ($t(43) = 33.74, p < .001$).

In addition, a word stimulus set was created consisting of 10 masculine (i.e., hoisting crane, tractor, race car, toy body shop, tools, rugby ball, power shovel, fire truck, pirate costume, helicopter) and 10 feminine toys (i.e., doll clothes, Barbie doll, doll, doll pram, toy kitchen, jewelry, doll house, princess costume, tea set, hoops). We only selected toys that were clearly identified as stereotypical masculine or feminine in previous research (Blakemore & Centers, 2005; Endendijk et al., 2013, 2014).

2.3.1.2. Task design. We adjusted the impression formation task of Li et al. (2016) to include pictures of boys and girls and toy words. Participants were instructed to form impressions based on the words paired with the faces based on their gut reactions, while giving evaluations on a 3-point rating scale (1 = positive; 2 = neutral; 3 = negative) on a button box. Ratings for each trial were recoded into −1 (negative), 0 (neutral), and +1 (positive), and averaged across trials. Use of 3-point rating scales is common practice in previous fMRI research (Boz, Spencer, & Montoya, 2018; Jacques, Dolcos, & Cabeza, 2010; Ochsner et al., 2004; Rasch et al., 2010). After the instructions, participants performed 12 practice trials prior to the event-related fMRI task. Across two runs, 20 unique face-word combinations (5 boy faces paired with masculine toys, 5 girl faces paired with masculine toys, 5 boy faces paired with feminine toys, and 5 girl faces paired with feminine toys) were presented 4 times each pseudorandomly (i.e., the same toy could not be paired with the same face more than once, first all 20 faces had to be presented, before another run started). This resulted in a total of 160 trials, with each face being paired with 4 masculine and 4 feminine toys. This paradigm thus yielded a 2 (Sex: child: Boy vs. Girl) × 2 (Toy type: Masculine vs. Feminine) within-subject design.

The stimulus face (size: 80 mm by 80 mm) was superimposed on a grey background (191;191;191), with the toy word (Courier new, point 3) presented in a random fashion to introduce jitter into the fMRI time-series to create inter-trial intervals of either 500 ms, 3000 ms, 5500 ms, or 2500 ms for a varying number of times (2, 3, or 4) and were interspersed in a random fashion to introduce jitter into the fMRI time-series to create inter-trial intervals of either 500 ms, 3000 ms, 5500 ms, or 8000 ms. Stimulus presentation, timing, and measurement of behavioral response time and accuracy were controlled by E-Prime (v 2.0) software (Schneider, Eschman, & Zuccolotto, 2002).

2.3.1.3. Scanning parameters. Scanning was performed on a 3 T Philips Achieva MRI scanner (Philips Medical Systems, Best, The Netherlands). Before the functional scans, a high resolution anatomical T1-weighted scan with the following parameters was obtained for co-registration and normalization purposes: 3.8 ms echo time, 8.4 ms repetition time, 288 × 288 × 175 mm field of view, 175 sagittal slices, flip angle of 80°, voxelsize 1.0 mm isotropic. Blood oxygen level dependent (BOLD) response was measured with functional T2*-weighted axial whole-brain images, of which 490 were obtained throughout the task. The 2D-EPI-SENSE sequence had the following parameters: echo time 24 ms, repetition time 1.01 s, 220 × 127.5 × 220 mm field of view, 51 slices.

Table 1

<table>
<thead>
<tr>
<th>Dimension</th>
<th>M(σD)</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mothers’ age</td>
<td>37.13 (5.29)</td>
<td>27–45</td>
</tr>
<tr>
<td>% highly educated</td>
<td>91%</td>
<td></td>
</tr>
<tr>
<td>% Dutch-Caucasian ethnicity</td>
<td>96%</td>
<td></td>
</tr>
<tr>
<td>Marital status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Married/registered partnership</td>
<td>64%</td>
<td></td>
</tr>
<tr>
<td>Single-mother</td>
<td>18%</td>
<td></td>
</tr>
<tr>
<td>Number of children</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>22%</td>
<td>1-4</td>
</tr>
<tr>
<td>2</td>
<td>74%</td>
<td>5-7</td>
</tr>
<tr>
<td>3</td>
<td>4%</td>
<td>8-10</td>
</tr>
<tr>
<td>Sibling gender composition</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Boys only</td>
<td>44%</td>
<td></td>
</tr>
<tr>
<td>Girls only</td>
<td>22%</td>
<td></td>
</tr>
<tr>
<td>Boy(s) &amp; girl(s)</td>
<td>35%</td>
<td></td>
</tr>
<tr>
<td>Age child 1</td>
<td>6.00 (2.22)</td>
<td>3-10</td>
</tr>
<tr>
<td>Age child 2</td>
<td>3.83 (2.07)</td>
<td>1-7</td>
</tr>
<tr>
<td>Age child 3</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Percentage of mothers with a child aged...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-1</td>
<td>4%</td>
<td></td>
</tr>
<tr>
<td>2-3</td>
<td>35%</td>
<td></td>
</tr>
<tr>
<td>4-5</td>
<td>74%</td>
<td></td>
</tr>
<tr>
<td>&gt; 5</td>
<td>44%</td>
<td></td>
</tr>
</tbody>
</table>

* a higher vocational or university level.

*Table 1: Demographic Characteristics of Sample.*
flip angle of 65°, voxel size 2.5 mm isotropic, SENSE-factor R = 3.0 (anterior-posterior).

2.3.2. Gender stereotypes about toys

Mothers completed a computer task to assess gender stereotypes about toys. The validity of this task to assess gender stereotypes in parents has been demonstrated (Endendijk et al., 2013). Participants were asked to divide toys between two children as quickly as possible, by means of pressing one of two keys on the keyboard that were assigned to each child. The two children (full-color) were presented constantly in the upper left and right hand corners of the monitor. Each full-color toy was presented in the middle of the screen until the participant pressed the response key. Trials were separated by an inter-stimulus interval of 500 ms.

The task started with a practice block (20 trials) in which red and blue presents had to be divided between two children of unknown gender, followed by two stereotype-congruent blocks and two stereotype-incongruent blocks (68 trials in each block). In the congruent blocks, mothers were instructed to assign stereotypically feminine toys (e.g., doll) to a girl and stereotypically masculine toys (e.g., car) to a boy. In the incongruent blocks, mothers were instructed to assign stereotypically feminine toys to a boy and stereotypically masculine toys to a girl. The order of the experimental blocks was: congruent-incongruent-congruent-incongruent. In this way, mothers made a switch between a congruent and incongruent block, and between an incongruent and congruent block. Participants were given a rest period between each block of self-determined length (instructions for the next block were also provided), and the entire task lasted between 5–10 min, depending on the length of the rest periods. The improved scoring algorithm by Greenwald, Nosek, and Benaji (2003) for the IAT was used to determine the level of gender stereotypes of the parent. A high positive score represented more di

2.3.3. Gender stereotypes about children’s behavior

The Child Rearing Sex-Role Attitude Scale (CRSAS, Freeman, 2007) was used to assess mothers’ stereotypes about boys’ and girls’ behavior and the way they should be parented. The questionnaire consisted of 19 items that were completed on a 5-point scale ranging from 0 (strongly agree) to 4 (strongly disagree). The questionnaire was designed in a way that the items concerned similar statements about boys and girls. For example: “Boys who exhibit ‘sissy’ behavior will never be well adjusted” and “Girls who are ‘tomboys’ will never be well adjusted”. Scores were averaged in one composite score (Cronbach’s alpha = .75), with higher scores indicating stronger stereotyped expectations about the behavior and parenting of boys and girls.

2.3.4. Internal motivation to respond without prejudice

Participants completed the IMS subscale of the motivation to respond without prejudice scale (Plant & Devine, 1998). The scale has 5 items (e.g., “Because of my personal values, I believe that having stereotypes about boys and girls is wrong”), all of which are measured on a 1- (strongly disagree) to 9-point (strongly agree) scale. The scores were averaged in one composite score (Cronbach’s alpha = .67), with higher scores indicating more motivation.

2.4. Analyses

For the statistical analyses of the behavioral data and brain-behavior associations, SPSS 23 (IBM analytics) was used with a significance level of α = 0.05.

2.4.1. Behavioral data

Evaluation ratings and reaction times (RTs) during the impression formation task were cleaned by removing trials in which RTs < 300 ms (11 trials in total). All variables were checked for normality and outliers (values that were more than 3.29 SD away from the group mean; Tabachnick & Fidell, 2012). We performed an ANOVA with Child sex and Toy type as within-subjects factors to examine whether mothers’ evaluated boys and girls differently when they were paired with either stereotype congruent or incongruent toys. We then separately added the gender-stereotype measures and internal motivation to respond without prejudice as covariates to this ANOVA to test whether the covariates affected the evaluation of the boys and girls in the impression formation task. Last, we computed correlations between the behavioral measures.

2.4.2. fMRI data

Scanning parameters are reported in the supplementary online material. Preprocessing and subsequent analyses were performed with SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Functional scans were motion corrected after which the anatomical scan was then coregistered to the mean functional scan. Subsequently, using unified segmentation, the structural scan was segmented and normalization parameters were estimated. Using these normalization parameters, all volumes were normalized to a standard brain template (MNI) and were resliced at 2 mm isotropic voxel size. Smoothing with a 6 mm full width at half maximum Gaussian kernel was applied to the normalized functional volumes. Next, a general linear model (GLM) was applied to the data to investigate the effects of stimulus conditions. Neural responses to the different face-word combinations were modeled using a 3 s boxcar function convolved with a hemodynamic response function (hrf) as implemented in the SPM12 software. Additional regressors of no interest which are entered into the analyses to reduce unexplained variance in the data include realignment parameters, a discrete cosine transform high-pass filter with a cutoff of 128 s and an hrf-convolved onset of the button press by which the participant rated the stimuli. BOLD responses were adjusted to the onset of the button press to guarantee attention of the participants to the stimuli.

The contrast maps of the different face-word conditions vs baseline were entered in a second-level factorial ANOVA, with stimulus type (boy picture-masculine toy; boy picture-feminine toy; girl picture-masculine toy; girl picture-feminine toy) as within-subjects factors. F-tests were conducted for the effect of child sex and toy type and their interaction. T-tests were performed to examine difference in (de)activations of stereotype incongruent (boy face-masculine toy; girl face-feminine toy) vs congruent stimuli (boy face-feminine toy; girl face-masculine toy). To control for multiple comparisons in the whole-brain analyses a threshold was set at p < 0.05 [family-wise error (FWE) corrected]. In addition, small volume corrections (SVC; p < 0.05 FWE) were applied for the predefined regions of the interest (ROIs). The ACC ROI was based on the automated anatomical labeling (AAL) template (Tzourio-Mazoyer et al., 2002). The temporoparietal junction (TPJ), dIPFC, dmPFC are not included in the AAL template as separate masks and were therefore derived from previous empirical papers. The mask for the TPJ was based on Cloutier et al. (2011) and consisted of 8 mm spheres around MNI coordinates −54, −56, 16 and 54, −51, 18. The mask for the dIPFC was based on Cloutier et al. (2011) and Li et al. (2016) and consisted of an 8 mm sphere around MNI coordinates 2, 54, 30. The mask for the dIPFC was based on Helman et al. (2014); Knutson et al. (2007) and Cattaneo et al. (2011) and consisted of 8 mm spheres around MNI coordinates −51, 16, 47 and 38, 22, 45. Contrast maps of the second-level analyses are publicly available in Neurovault (via the following link: added after publication).

2.4.3. Exploratory brain-behavior associations

For all the predefined anatomical ROIs that showed significant effects in the whole brain analyses, we extracted percent signal change using MarsBaR (Brett, Anton, Valabregue, & Poline, 2002) to further
specify and visualize the effects. We also ran exploratory correlational analyses to investigate whether any of the gender stereotype measures were associated with increased brain activity to incongruent versus congruent trials. These correlations were only computed for the ROIs for which significant effects (or trends) of congruence were found.

3. Results

3.1. Behavioral data

Table 2 shows descriptive statistics and correlations for behavioral data.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.41 (0.38)</td>
<td>0.41 (0.40)</td>
<td>0.44 (0.42)</td>
<td>0.39 (0.38)</td>
<td>0.02 (0.33)</td>
<td>0.17 (0.40)</td>
<td>6.55 (1.37)</td>
</tr>
</tbody>
</table>

Note. Evaluations were assessed in the fMRI task. Gender stereotypes about children’s behavior and internal motivation to respond without prejudice were assessed with self-report questionnaires. Gender stereotypes about toys were assessed with a computerized implicit association task.

\* \( p < .05 \)

\** \( p < .01 \)

Table 3 and Fig. 3 display the brain regions that became activated as a function of Sex, Toy type, and Congruence of the stimuli. A significant interactive effect between child sex and toy type was found for left-TPJ activity. To interpret this interaction a 2 (Sex: boy vs. girl) × 2 (Toy type: masculine vs. feminine toy) RM-ANOVA was performed on extracted left-TPJ activity. A main effect of child sex was found, \( F(1, 22) = 5.62, p = .027 \), which was subsumed by a significant two-way interaction between child sex and toy type, \( F(1, 22) = 5.09, p = .034 \), \( partial \eta^2 = .19 \) (see Fig. 3A). Post-hoc t-tests indicated a trend toward a greater left-TPJ response for pictures of boys paired with feminine toys (\( M = 0.07, SD = 0.49 \)) compared to pictures of boys paired with masculine toys (\( M = -0.06, SD = 0.59 \), \( t (23) = 1.75, p = .095 \)). This difference was not found for the pictures of girls, \( t(23) = 0.88, p = .39 \).

A significant effect of congruence on dmPFC activity and a trend
toward significance on ACC activity suggested increased activity in these areas in response to stereotype-incongruent stimuli (e.g., picture of a girl paired with the toy “car”) compared to stereotype-congruent stimuli (e.g., picture of a boy paired with the toy “car”) (Fig. 3B). Furthermore, pictures of girls elicited a greater response in the primary visual cortex (calcarine gyrus) compared to pictures of boys.

3.3. Exploratory brain-behavior associations

Exploratory correlational analyses (for left-TPJ, ACC, dmPFC) showed that a greater ACC response to incongruent than congruent trials was associated with stronger gender stereotypes about toys (Fig. 4; left-ACC: \( r = .44, p = .034 \); right-ACC: \( r = .50, p = .015 \)). A similar trend was found for a greater dmPFC response to incongruent than congruent trials being associated with stronger gender stereotypes about toys (\( r = .37, p = .081 \)). A greater left-TPJ response to incongruent than congruent trials was specifically associated with stronger stereotypes about boys’ and girls’ behavior and the way they should be parented (\( r = .47, p = .023 \); Fig. 5). Because differential left-TPJ responses to incongruence were more pronounced for boy pictures than for girl pictures, we also examined whether the correlation between TPJ activity and gender stereotypes was different for boy (\( r = .39, p = .07 \)) and girl pictures (\( r = .16, p = .46 \)), which was not the case (\( r = .68, p = .25 \)). All correlational analyses were uncorrected for multiple comparisons and should thus be interpreted with caution. Results for all brain-behavior correlations were the same whether we controlled for sibling gender composition or not.

In order to replicate the findings of Li et al. (2016) with regard to internal motivation to respond without prejudice (IMS) we conducted a regression analysis with IMS as predictor and dmPFC activity to incongruent versus congruent trials as outcome. IMS was not significantly related to dmPFC activity, \( b = 0.075, SE = 0.046, t(22) = 1.64, p = 0.12 \) (whether we controlled for other gender stereotype measures or not).

4. Discussion

This study examined mothers’ neural responses to pictures of boys and girls paired with gender-typical or gender-atypical objects. Analyses showed that mothers showed stronger ACC activity in response to stereotype-incongruent stimuli compared to stereotype-congruent stimuli. Additionally, pictures of girls elicited a greater response in the primary visual cortex (calcarine gyrus) compared to pictures of boys. Exploratory correlational analyses revealed that a greater ACC response to incongruent than congruent trials was associated with stronger gender stereotypes about toys. A similar trend was found for a greater dmPFC response to incongruent than congruent trials being associated with stronger gender stereotypes about toys. A greater left-TPJ response to incongruent than congruent trials was specifically associated with stronger stereotypes about boys’ and girls’ behavior and the way they should be parented. Because differential left-TPJ responses to incongruence were more pronounced for boy pictures than for girl pictures, the correlation between left-TPJ activity and gender stereotypes was different for boy and girl pictures. All correlational analyses were uncorrected for multiple comparisons and should thus be interpreted with caution. Results for all brain-behavior correlations were the same whether we controlled for sibling gender composition or not.

In order to replicate the findings of Li et al. (2016) with regard to internal motivation to respond without prejudice (IMS) we conducted a regression analysis with IMS as predictor and dmPFC activity to incongruent versus congruent trials as outcome. IMS was not significantly related to dmPFC activity, \( b = 0.075, SE = 0.046, t(22) = 1.64, p = 0.12 \) (whether we controlled for other gender stereotype measures or not).

Table 3

<table>
<thead>
<tr>
<th>Experimental effect</th>
<th>Peak voxel location</th>
<th>t/F-value</th>
<th>Cluster size</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>x y z</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full factorial</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F-test: sex child x toy type</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TPJ</td>
<td>L −52 −54</td>
<td>7.14</td>
<td>7</td>
<td>0.029**</td>
</tr>
<tr>
<td>T-test: child sex (girl &gt; boy)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Calcarine gyrus</td>
<td>L −14 −96 −24</td>
<td>6.08</td>
<td>20</td>
<td>0.001**</td>
</tr>
<tr>
<td>T-test: congruence condition (IC &gt; C)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACC</td>
<td>0 40 24</td>
<td>3.74</td>
<td>17</td>
<td>0.069**</td>
</tr>
<tr>
<td>dmPFC</td>
<td>2 56 26</td>
<td>3.59</td>
<td>13</td>
<td>0.012**</td>
</tr>
</tbody>
</table>

Note. TPJ, temporal parietal junction; ACC, anterior cingulate cortex; dmPFC, dorsomedial prefrontal cortex; R, right; L, left.

* whole brain FWE corrected at cluster level.
** small volume FWE corrected at cluster level.

---

**Fig. 3.** A) Coronal slice with corresponding Y-coordinates (MNI) from the F-map of the interaction between child sex and toy type overlaid onto a standard anatomical template. On the right panel the extracted parameter estimates from the left-TPJ are plotted for all conditions. B) Axial slices with corresponding Z-coordinates (MNI) from the T-map of neural activation for incongruent versus congruent stimuli on which activation of the dmPFC and ACC are visible. Accompanying statistics are described in the text. All statistical maps are thresholded at \( p = 0.001 \) uncorrected, for illustration purposes only.
and girls who confirmed or violated social expectations regarding toy preferences. In addition, we explored whether neural sensitivity to targets violating gendered expectations interacted with mothers’ gender stereotypes and beliefs. We found that exposure to child targets that violated gender expectations for toy preferences increased neural activation in the ACC and dmPFC, but not in the dIPFC. In addition, left-TPJ activity specifically increased in response to pictures of boys who violated expectations regarding toy preference. Furthermore, increased neural responses in the ACC, dmPFC, and left-TPJ to targets violating gender expectations was associated with more stereotyped expectations about boys and girls.

Our finding that child targets violating gender expectations for toy preferences increased neural activation in the ACC and dmPFC, and (left)-TPJ is in line with neural models of stereotypes (Amodio, 2014; Stanley et al., 2006) and previous fMRI research on violations of social expectations (Cloutier et al., 2011; Hehman et al., 2014; Li et al., 2016). More specifically, Li et al. (2016) demonstrated increased dmPFC activity in response to adult targets that violated racial stereotypes, whereas Hehman et al. (2014) found stronger activation in the dmPFC and ACC in response to adult targets that violated racial stereotypes about emotions. In addition, Cloutier et al. (2011) showed that the TPJ was recruited in response to Republican and Democrat politicians who violated political views. Our findings extend previous research by showing that these areas are not only involved in responses to adult targets that violate social expectations, but also in parents’ processing of child targets that confirm or violate gender expectations about toys. Furthermore, together the body of research shows that the ACC, dmPFC, and TPJ might be part of a general stereotyping network, since these brain areas become activated regardless of the stereotype domain that is violated (i.e., gender, race, politics).

These findings also provide an important contribution for research on parenting by pointing to several possible unconscious processes that might underlie parents’ negative responses to boys’ and girls’ behavior that violates gender-role expectations. Involvement of the ACC when parents are exposed to children who violate gender expectations might suggest that parents experience conflict when a child’s behavior does not match social expectations about gender (Stanley et al., 2008). There is ample evidence for the role of the ACC in conflict monitoring, for example in tasks the require overriding of prepotent responses or the selection of a response from equally permissible responses (for reviews, see Alexander & Brown, 2011; Botvinick et al., 2004). Yet, the lack of differences in mothers’ ratings of children who violated versus confirmed gender expectations, might also suggest that the ACC has successfully activated areas in the prefrontal cortex responsible for top-down behavioral control (Amodio, 2014). The dmPFC might be implicated in this regard as it showed increased activation in response to child targets that violated gender expectations. The dmPFC has indeed been associated with inhibition of gender-stereotyped behavior (Cattaneo et al., 2011) and with monitoring behavior and control of response selections (for a review, see Bushworth et al., 2007). However, the role of the dmPFC is not completely clear, as involvement of the dmPFC could also indicate that parents are updating their repository of social knowledge about gender over time when they are exposed to children who violate gender expectations (Van Overwalle, 2009). If this is the case, frequent exposure to social targets that violate social expectations, might lead to a more diverse and nuanced repository of knowledge about certain social groups (Gawronski & Bodenhausen, 2006). Involvement of the dmPFC might also suggest that parents are forming an impression of a child’s traits, preferences, and mental states based on gender-stereotyped-expectancies (Cloutier et al., 2011; Li

Fig. 4. The ACC activity in response to gender incongruence varies as a function of Mothers’ gender stereotypes about toys. A) Correlation between ACC response to gender-incongruent versus congruent stimuli and mothers’ gender stereotypes about toys. B) The figure represents the effect of stimulus congruence on ACC activity as a function of mothers’ gender stereotypes. Median split performed on mothers gender stereotypes for visualization purposes only.

Fig. 5. Left-TPJ activity in response to gender incongruence varies as a function of Mothers’ gender stereotypes about parenting. A) Correlation between left-TPJ response to gender-incongruent versus congruent stimuli and mothers’ gender stereotypes about parenting boys and girls. B) The figure represents the effect of stimulus congruence on left-TPJ activity as a function of mothers’ gender stereotypes about parenting boys and girls. Median split performed on mothers gender stereotypes for visualization purposes only.
et al., 2016). Last, increased TPJ responses to boy targets that violated social expectations about toy preferences could suggest that parents are trying to understand the goals or intentions behind the toy preferences shown by the boys (Amodio, 2014; Van Overwalle, 2009). The TPJ has indeed been linked to reasoning about others’ mental states (for a review, see Saxe, 2006; Young et al., 2010).

As hypothesized, we only found increased left-TPJ activity in response to pictures of boys who violated expectations regarding toy preference, and not to girls who violated these expectations. This finding might reflect a more restrictive nature of stereotypes about male roles than stereotypes about female roles (Hort, Fagot, & Leinbach, 1990; Leaper, 2000). Indeed, parents have been found to respond more negatively when boys behave in a way that violates gender-role expectations, compared to girls behaving in the same way (Kane, 2006; Sandnabba & Ahlberg, 1999). For example, mothers may think it is more appropriate and expected for girls to play soccer than it is for boys to do ballet, which could explain mothers’ increased neural response to pictures of boys, and not girls, who violated expectations regarding toy preferences. Relatedly, parents and other adults have been found to be more concerned with socializing boys to show gender-typical behavior than they are with girls (Egan & Perry, 2001; Thomas & Blakemore, 2013).

Unexpectedly, we did not find differential activation of the dlPFC in response to boys and girls who violated versus confirmed expectations about toy preferences. This finding is not in line with a previous study showing increased dlPFC activation in response to adult men and women who violated gender stereotypes about occupations (Quadflieg et al., 2011). The difference between these studies could be due to differences in target age and parental status of the participants. It has been argued that children may be less likely than adults to encounter backlash for violating gendered expectations (Sullivan, Moss-Racusin, Lopez, & Williams, 2018), because children are generally punished less severely for violations of cultural norms (Laney, 2015). In addition, stereotype violations by children do not disrupt existing power hierarchies (Rudman, Moss-Racusin, Phelan, & Nauts, 2012). Thus, stereotype violations might be accepted more from children than from adults, especially by parents who know that children are still learning about gender roles. Yet, it remains to be tested whether these processes translate into differential engagement of the dlPFC to regulate behavior and inhibit stereotype-based responses when people encounter stereotype-violating children versus stereotype-violating adults.

Alternatively, it has been suggested that the dlPFC might be recruited in particular in situations that require cognitive control to inhibit stereotype-based responses (Amodio, 2014; Stanley et al., 2008). Our task, in which mothers were asked to evaluate the boys and girls in the pictures based on their gut reactions, might not have required or elicited much cognitive control. Other tasks, such as IATs might require more cognitive control as prepotent responses need to be inhibited. Indeed, fMRI studies using IATs more consistently demonstrate involvement of the dlPFC (Amodio, 2014; Cattaneo et al., 2011; Knutson et al., 2007), than studies using paradigms based on violations of social expectations, such as our task (Cloutier et al., 2011; Hehman et al., 2014; Li et al., 2016; Quadflieg et al., 2011). The former paradigm might represent a more implicit measure of stereotype expectations, whereas the latter paradigm might represent a more explicit measure of stereotyped expectations. With explicit stereotype measures people are generally more aware that stereotyped expectations are assessed than with implicit stereotype measures (Greenwald, Poehlman, Uhlmann, & Banaji, 2009). Future research could directly test differences in the neural basis of implicit and explicit stereotyped expectations, by using an IAT paradigm next to a violations-of-social-expectations paradigm. In both tasks it is possible to use the same stimuli, in order to make direct comparisons.

We also found a non-hypothesized effect of target sex on activation in the primary visual cortex, with pictures of girls eliciting more activation in the area of the calcarine sulcus. This finding could reflect enhanced processing of in-group targets, as mothers may likely identify with the same social group as the girl targets (Ito & Urland, 2003). This finding could also indicate that male and female faces are processed differently on the basis of low-level stimulus properties (Kaul, Rees, & Ishai, 2011). Indeed, a previous fMRI study showed that the early visual cortex is involved with gender discrimination and encoding of faces, regardless of participant gender or sexual orientation (Kaul et al., 2011). Because of the evolutionary importance of gender discrimination (i.e., sexual selection, mating) and its fundamental nature in face processing (Ng, Tay, & Goi, 2015), it seems likely that some gender-specific features of a face (e.g., hair-length, forehead size) are processed in earlier visual areas around the calcarine sulcus. Future studies could test whether activation in the area of the calcarine gyrus reflects enhanced processing of in-group targets or a more general gender discrimination in face processing. For instance, the current study design could be expanded to include fathers as well as mothers.

Not all mothers responded the same to child targets that violated versus confirmed gender stereotypes, showing within-group variation and potential existence of important individual differences in responses. Previous studies with similar samples sizes as ours, and therefore limited with respect to testing individual differences, have also found meaningful associations between people’s stereotypes and neural responses to stereotypical stimuli (e.g., Hehman et al., 2014). However, replication with larger samples is needed, as some of these findings might in retrospect be considered false positives. We found that, in particular, more stereotyped expectations about boys and girls were associated with increased neural responses in the ACC, dmPFC, and left-TPJ to targets that violated gender expectations. These findings are similar to a study on racial stereotypes, demonstrating that adults with strong racial stereotypes about emotions showed stronger neural activation (in the dlPFC) to individuals who violated racial stereotypes (Hehman et al., 2014). An EEG study also demonstrated that increased brain activity to gender-incongruent toy-child pairings was associated with stronger gender stereotypes in mothers (Endendijk, Groeneveld et al., 2018, 2018b). Our findings could suggest that mothers who have stereotyped expectations about boys’ and girls’ toy preferences, behavior, and the way they should be parented, might experience increased conflict, as reflected by increased ACC activity, when a child’s behavior does not match their expectations about gender (Stanley et al., 2008). Subsequently, the heightened conflict signalling in the ACC might increasingly engage dlPFC regions to control gender-stereotyped responses (Amodio, 2014; Cattaneo et al., 2011; Rushworth et al., 2007). These mothers’ might also have more difficulty with forming an impression of a child’s traits, preferences, and mental states (Cloutier et al., 2011; Li et al., 2016) and with understanding the goals or intentions behind the stereotype-violating behavior of the child (Amodio, 2014; Van Overwalle, 2009; Saxe, 2006; Young et al., 2010), as reflected respectively by increased dmPFC and TPJ activity. Future studies could examine how this neural processing of boys and girls who violate gender expectations translates to mothers’ actual behavior with their own sons and daughters.

Unexpectedly, we did not find an association between mothers’ self-reported motivation to act without prejudice about gender and neural responses to child targets that violated gender expectations. Yet, mothers with lower motivation evaluated child targets that violated gender expectations less positive than child targets that confirmed social expectations. Apparently, mothers’ internal motivation to respond without prejudice was not associated with diminished neural sensitivity to stereotype-incongruence. A previous study did demonstrate that internal motivation to respond without prejudice attenuated the greater dmPFC activity to Black and White adult faces who violated racial attitudes, compared to individuals who confirmed racial attitudes (Li et al., 2016). It is possible that internal motivation to respond without prejudice might be more relevant in the context of racial stereotypes than in the context of gender stereotypes. Social norms discouraging expressions of gender-based prejudice may not be as compelling or
consistent as the norms discouraging race-based prejudice (Eberhardt & Fiske, 1994). Relatedly, people generally have quite positive attitudes and stereotypes of women, which may reduce the strength of internal motivation to avoid gender-based prejudice (Klonis, Plant, & Devine, 2005). It is also possible that the lack of associations found between internal motivation to respond without prejudice and neural responses might be due to the low internal consistency for this measure in the current study, or due to a lack of power.

The results need to be interpreted while considering the limitations of this study. First, we examined mothers with children in a wide age range, including infants, preschoolers, and school-aged children, which may have affected our results. However, in the current study, controlling for child age did not change our findings and children on the pictures in the fMRI task were in the same age range as mothers’ own children.

Second, across the impression formation task each target was paired randomly with 4 stereotype-confirming and 4 stereotype-violating descriptions. Therefore, impression updating may have occurred, which has been found to be specifically associated with dmPFC activity when people are presented with information that is inconsistent with previous information about a target (Ferrari, Vecchi, Todorov, & Cattaneo, 2016; Mende-Siedlecki, Cai, & Todorov, 2012). However, impression updating is generally studied with designs in which the same face is paired consecutively with different descriptions. Yet, in the current study, due to the pseudorandom presentation of the stimuli, the same faces were never presented consecutively and all 20 unique faces had to be presented once before they were presented for the second time. Therefore, impression updating may have only played a marginal role in the recruitment of the dmPFC in response to child targets that violated gender stereotypes.

Third, we used a 3-point rating scale (negative to positive) to assess mothers’ evaluation of children whose toy preferences violated or confirmed gender stereotypes. Although such rating scales have been used successfully in previous fMRI research (e.g., Bos et al., 2018; Jacques et al., 2010; Ochsner et al., 2004; Rasch et al., 2010), it was probably not sensitive enough to detect differences in mothers’ evaluation of stereotype congruent and incongruent face-toy combinations. Rating scales with more response options (e.g., 5 or more) might yield more variability in participants evaluation of targets in impression formation tasks (Li et al., 2016). In addition, our homogeneous sample in terms of education (high) and ethnicity (Caucasian) might provide another explanation for the lack of differences in the evaluation of children who violated versus confirmed gender stereotypes. There is evidence that higher education is associated with more egalitarian gender stereotypes (e.g., Dodson & Borders, 2006), and possibly also of greater acceptance of behavior that violates gender stereotypes.

Fourth, because of our modest sample size we were not able to examine differences between mothers with sons, mothers with daughters, and mothers with both sons and daughters. A direction for future research is to examine predictive relationships between functional brain activity in response to gendered stimuli and observed gendered parenting with boys and girls, in diverse samples with enough power to detect individual differences. Such studies could increase our understanding of specific brain areas underlying parents’ negative responses to children’s behavior that violates gender-role expectations. It is also important to include fathers in these studies, because fathers were found to be more concerned with their children, and especially their sons, conforming to gender expectations (Kane, 2006).

To conclude, this study highlights the importance of examining individual differences in neural responses to stimuli that violate versus confirm social expectations. Individual differences in mothers’ stereotypical expectations about boys and girls appear to be associated with the involvement of the ACC, dmPFC, and left-TPJ, brain regions believed to be central to social stereotyping. More specifically, mothers with more egalitarian expectations about boys’ and girls’ toy preferences, behavior, and the way they should be parented, may be less sensitive at a neural level to stereotype-incongruent behavior in boys and girls. Our findings also point to several unconscious processes that might underlie parents’ negative responses to children’s behavior that violates gender-role expectations, and boys’ gender-role inconsistent behavior in particular. First, mothers might view boys as members of a generic social category instead of unique individuals. Second, mothers might experience conflict when a child’s behavior does not match parents’ social expectations about boys and girls behavior. Yet, a neural control mechanism is also engaged to inhibit gender-stereotyped responses. Third, mothers might have more difficulty with forming an impression of a child’s goals and intentions. However, it remains to be examined how these neural mechanisms translate to negative responses to actual children who behave in a way that is not consistent with gender stereotypes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was supported by a grant from the Netherlands Society of Scientific Research (451-14-015) to PAB.

References
