



Measuring the neural correlates of the violation of social expectations: A comparison of two experimental tasks

Christel M. Portengen, Rens Huffmeijer, Anneloes L. van Baar & Joyce J. Endendijk

To cite this article: Christel M. Portengen, Rens Huffmeijer, Anneloes L. van Baar & Joyce J. Endendijk (2022) Measuring the neural correlates of the violation of social expectations: A comparison of two experimental tasks, Social Neuroscience, 17:1, 58-72, DOI: [10.1080/17470919.2022.2032327](https://doi.org/10.1080/17470919.2022.2032327)

To link to this article: <https://doi.org/10.1080/17470919.2022.2032327>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



[View supplementary material](#)



Published online: 09 Feb 2022.



[Submit your article to this journal](#)



Article views: 84



[View related articles](#)



[View Crossmark data](#)

RESEARCH ARTICLE



Measuring the neural correlates of the violation of social expectations: A comparison of two experimental tasks

Christel M. Portengen^a, Rens Huffmeijer^b, Anneloes L. van Baar^a and Joyce J. Endendijk^a

^aChild and Adolescent Studies, Clinical Child and Family Studies, Utrecht University, Utrecht, The Netherlands; ^bInstitute of Education and Child Studies, Leiden University, Leiden, The Netherlands

ABSTRACT

Evidence exists that people's brains respond differently to stimuli that violate social expectations. However, there are inconsistencies between studies in the event-related potentials (ERP) on which differential brain responses are found, as well as in the direction of the differences. Therefore, the current paper examined which of the two most frequently used tasks, the Impression Formation Task (IFT) or Implicit Association Test (IAT), provided more robust ERP components in response to the violation of gendered expectations. Both IFT and IAT paradigms were administered in a counter-balanced way among 25 young adults (age 22–31, 56% male), while brain activity was assessed with electroencephalography. The IFT and IAT specifically measured the violation of gendered expectations with regard to toy preferences and behavioral tendencies of young children. The results showed that both tasks were able to elicit relevant ERP components. Yet, the IFT evoked ERP effects of the violation of gendered expectations on all but one of the selected ERP components; the P1, N1, and LPP. The IAT only elicited different P3 amplitudes when expectations were violated. We recommend the use of IFT paradigms when studying neural processes underlying the violation of social expectations.

ARTICLE HISTORY

Received 11 May 2021
Revised 17 December 2021
Published Online 10
February 2022

KEYWORDS

Violations of social expectations; gender; event-related potentials; implicit association test; impression formation task

A large body of literature demonstrates that people respond negatively to violations of social expectations (Endendijk et al., 2014; Friedman et al., 2007; Kane, 2006; Sandnabba & Ahlberg, 1999). This also holds for gendered expectations. For instance, parents believe that children who do not adhere to traditional gendered expectations, will be less psychologically well adjusted than “typical” boys and girls later in life (Sandnabba & Ahlberg, 1999). Parents are also more likely to make negative evaluative comments about children's behavior that violates gendered expectations (e.g., “Boys don't play with dolls!”; Endendijk et al., 2014; Friedman et al., 2007). Not only parents, but also non-parental adults rate children who violate gendered expectations as less likeable (Sullivan et al., 2018). Moreover, women who have a successful career in traditionally male-dominated work environments, are perceived as more hostile and less likeable, which affects their overall performance ratings, salary, and job opportunities (Heilman et al., 2004). To understand why negative responses to violations of social expectations with regard to gender, race, or even age, occur, neuroscientists have tried to uncover how the human brain processes the violation of social expectations.

To examine the neural processes underlying peoples' negative reactions to violations of social expectations, researchers often relied on one of the following experimental paradigms: The Implicit Association Test (IAT) and the Impression Formation Task (IFT). Previous research demonstrated that both paradigms are able to elicit meaningful event-related potentials (ERPs), which are time-locked epochs of neural activation patterns that occur around the presentation of a stimulus that violates or confirms social expectations (Dickter & Gyurovski, 2012; Forbes et al., 2012; He et al., 2009; Healy et al., 2015; Hehman et al., 2014; Li et al., 2016; Williams & Themanson, 2011; Xiao et al., 2015). However, studies differ in the ERP components on which differential brain responses are found. In addition, the reasons for choosing one paradigm over the other are often not explicitly stated in previous studies. These issues might be attributed to the fact that we do not yet know which paradigm is better suited to capture processing of violations of social expectations. Therefore, the current study examines which experimental paradigm, the IAT or the IFT, elicits the most robust pattern of ERPs. We will study this in the context of people's brain responses to violations of gender expectations, since gender is one of the

CONTACT Joyce J. Endendijk  j.j.endendijk@uu.nl  Department of Clinical Child and Family Studies, Utrecht University, Utrecht, The Netherlands
 Supplemental data for this article can be accessed [here](#)

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

most salient social categories that people use to categorize preferences, occupations, and behaviors from birth onwards (Blakemore et al., 2008).

Both the IAT and the IFT are established tasks to measure responses to violations of social expectations in several domains. The IAT is a response latency task in which participants are asked to divide attributes (science, career, family, caring, male-typed toys, female-typed toys, male, female, boys, girls) across two categories (Greenwald et al., 2003, 2009). The task consists of two blocks that are congruent with social expectations, meaning that the attributes must be assigned to a category that confirms the culturally prescribed social expectations (e.g., “science” and “male” both have to be assigned to the same category). In addition, there are two incongruent blocks, in which attributes must be assigned to a category that violates social expectations (e.g., “family” and “male” have to be assigned to the same category). The IFT, on the other hand, generally requires less active participation of the participants, as they are asked to form impressions of people based on combinations of stimuli that either confirm social expectations (e.g., a picture of a boy with the word “soccer”) or violate expectations (e.g., a picture of a boy paired with the word “doll”; Li et al., 2016). In both tasks, brain activity can be compared between trials that confirm social expectations and trials that violate social expectations.

Although both tasks have been shown to elicit meaningful ERPs when they are administered while participants’ brain activity is recorded with electroencephalography (EEG), both tasks also come with their own caveats. First, the IAT is known to elicit preparatory brain activity due to the block design in which congruent and incongruent trials are separated. Due to this block design, participants are aware of the upcoming violation of social expectations in the incongruent blocks and confirmation of social expectations in the congruent blocks. This awareness has led to systematic (preparatory) differences in pre-stimulus baseline brain activity between congruent and incongruent trials in previous studies (Bidet-Caulet et al., 2012; Endendijk et al., 2019b; Healy et al., 2015). To correct for these systematic differences in pre-stimulus baseline brain activity, preprocessing of the EEG data needs to be adjusted. This adjusted preprocessing can result in reduced power, or in lower EEG frequencies being filtered out, making especially late-ERP components less reliable (Demiralp et al., 2001). Nonetheless, in previous EEG research using the IAT several relevant ERP components were elicited that were related to behavioral

assessments of stereotypes or stereotyped behavior (e.g., Endendijk et al., 2019b; Forbes et al., 2012; Healy et al., 2015).

In the IFT, the IAT-specific issue regarding preparatory brain activity does not occur, since congruent and incongruent trials are presented randomly throughout each block. However, only passively viewing combinations of stimuli may lead to reduced attention to the task in participants. Especially with the large number of trials needed for EEG data, participants may lose their focus after a while in a passive viewing paradigm such as the IFT. To counteract the expected loss of focus, participants are often instructed to evaluate the stimuli by pressing a button. It, however, remains important to take these limitations into account when using the IFT in an EEG study on violations of social expectations.

The ERP components that are most often studied in the context of violations to social expectations are the P1 (He et al., 2009; Liu et al., 2017), N1 (Dickter & Gyurovski, 2012; Healy et al., 2015), P2 (Dickter & Gyurovski, 2012; He et al., 2009; Healy et al., 2015; Jerónimo et al., 2017; Yang et al., 2020), N2/Medial Frontal Negativity (MFN; Dickter & Gyurovski, 2012; Healy et al., 2015; Hilgard et al., 2014), P3 (Fabre et al., 2015; Healy et al., 2015; Siyanova-Chanturia et al., 2012), N400 (Proverbio et al., 2018, 2017; Siyanova-Chanturia et al., 2012; White et al., 2009; Yang et al., 2020; Yao & Wang, 2014), and Late Positive Potential (LPP; Liu et al., 2017; Osterhout et al., 1997; Rodríguez-Gómez et al., 2020; White et al., 2009; Yao & Wang, 2014). The first three ERP components (P1, N1, P2) have been associated with attentional processes; specifically with early and later stages of visual processing (P1, P2; Luck, 2014), and with visual discrimination processes (N1; Luck et al., 2000; Vogel & Luck, 2000). Regarding expectancy violations, female faces with an angry facial expression (i.e., expectation-incongruent) have been found to elicit higher P1 amplitudes than female faces with a happy facial expression (i.e., expectation-congruent) (assessed with a gender categorization task; Liu et al., 2017). The N1 has been found to show larger peaks during expectancy-violating than expectancy-confirming trials during a racial IFT (Dickter & Gyurovski, 2012) but not during an IAT measurement of stereotype violations (Healy et al., 2015). P2 amplitudes were found to be larger when a positive impression of a person is violated by presenting negative behaviors than when a negative impression of an individual is confirmed (assessed with an IFT; Jerónimo et al., 2017). Yet, P2 has also been found to be smaller when stereotypes about lower-status social groups (e.g., homeless people, drug addicts) are violated than when stereotypes are confirmed (assessed with IFT; Yang et al., 2020).

N2 amplitudes (or medial frontal negativity) are often associated with conflict monitoring or overcoming one's stereotyped responses in expectancy violation paradigms (Azizian et al., 2006). Expectancy violating racial stimuli yielded greater N2 amplitudes than expectancy-confirming racial stimuli during an IFT (Dickter & Gyurovski, 2012). In contrast, using an IAT, Healy et al. (2015) found increased N2 activity during stereotype-confirming trials compared to stereotype-violating trials, specifically when participants held medium stereotyped scores.

The P3 latency has been associated with attention to unexpected events (Polich, 2007). Not surprisingly, expectancy-violating stimuli have been found to elicit a larger P3 than expectancy-confirming stimuli in an IFT (Bartholow & Dickter, 2007). The N400 is thought to reflect the cognitive effort needed to integrate a stimulus into a given context, with larger peaks during expectancy-violating trials (e.g., "Women" followed by "Aggressive") compared to expectancy-confirming trials (e.g., "Women" followed by "Nurturing") during IFT paradigms (Proverbio et al., 2017; White et al., 2009; Yang et al., 2020). Finally, the LPP is elicited during tasks in which emotional stimuli are presented and shows a larger amplitude when these stimuli are more salient, for instance, when two stimuli of different modalities (e.g., pictures and words) but with similar valence (i.e., congruent) are combined (Spreckelmeyer et al., 2006). LPP was larger in response to sentences that violated gender expectations (i.e., "the doctor prepared herself for the operation") than to sentences that confirmed gender expectancies (i.e., "the doctor prepared himself for the operation"; Osterhout et al., 1997). On the other hand, angry male facial expressions (i.e., expectancy-confirming) elicited higher LPP amplitudes than happy male faces (i.e., expectancy-violating; Liu et al., 2017).

To summarize, both the violation and confirmation of expectations have evoked ERP components relevant for attentional processing, conflict monitoring, attention to unexpected events, stimulus-context integration, and evaluation of salience. However, there are differences between studies in the ERP components that are elicited by expectancy-violating versus expectancy-confirming stimuli, as well as in the direction of effects that is found (i.e., whether expectancy-violating stimuli elicit larger ERPs than expectancy-confirming stimuli, or the other way around). These between-study differences might be attributed to the different methods (IAT, IFT, categorization paradigm, sentences, words, pictures) that are applied in EEG studies.

In addition, individual differences in the way people process stimuli that violate or confirm social expectations may play a role in whether expectancy-violating or expectancy-confirming stimuli elicit larger ERP amplitudes (Canal et al., 2015; Endendijk et al., 2019b; Healy et al., 2015). For example, a person who strongly expects women to be emotional and men to be stoic, might react differently (on a behavioral and neural level) to a crying man than a person who expects men and women to be equally emotional. Therefore, it is important to consider people's stereotyped expectations and attitudes when examining neural responses to stimuli that violate social expectations. For instance, a study by Endendijk et al. (2019b) demonstrated that P3 and N2 activity elicited by stereotype-violating versus stereotype-confirming trials in an IAT were related to mothers' gender stereotypes. Similarly, Healy et al. (2015) found that N2 amplitudes were larger during stereotype-confirming trials in an IAT when people held medium stereotypes. Moreover, differences in LPP amplitudes to stereotype-confirming and stereotype-violating trials were related to participants' hostile sexism (Canal et al., 2015).

Ultimately, there is a large variability in the direction of effects and the ERP components on which effects are found when social expectations are violated, most likely because of the multiple ways to measure people's reactions to the violation of social expectations. Moreover, differences exist in the extent to which people hold different expectations with regard to social norms. The current study was designed to determine which one of two experimental paradigms (the IAT or the IFT) elicited the most robust differences in ERP components (P1, N1, P2, N2, P3, N400, LPP) between trials that confirmed (stereotype-congruent) versus violated (stereotype-incongruent) gender expectations. Robustness in this study was defined as the paradigm's ability to elicit meaningful (i.e., in line with previous research) and consistent ERP components across two types of stereotype-confirming and stereotype-violating stimuli types (i.e., violating/confirming gender expectations about toys and behavior). We specifically examined people's responses to the violation of gendered expectations by young children, since children still show very explicit forms of gendered behaviors in their toy play (i.e., playing with cars or dolls) and behaviors (i.e., openly crying or showing aggressive behaviors; Blakemore et al., 2008). Children who violate gendered expectations have evoked negative

reactions in both parents and non-parents (Endendijk et al., 2014; Sandnabba & Ahlberg, 1999; Sullivan et al., 2018).

Methods

Participants

Participants were 25 young adults (56% men) between the ages 22 and 31 ($M_{age} = 26.1$, $SD = 2.77$). Most participants were highly educated (i.e., 92% held a higher vocational degree or university degree). All participants were non-parents with normal or corrected-to-normal vision. Exclusion criteria were a neurological disease (e.g., Parkinson disease, multiple sclerosis), or a history of epileptic seizures, as these conditions are known to disturb EEG signals.

Procedures

Participants were recruited via the researchers' personal networks and university-related Facebook groups. Participants were unaware of the gendered nature of the study and were told to perform two tasks to see if they worked properly.

Participants were visited at home and they were asked to perform two tasks whilst undergoing simultaneous EEG examination. An IFT was administered, in which participants had to passively view child faces and toy/behavior words and rate face-word combinations on a scale from 1 to 9. In addition, participants were asked to perform a modified IAT in which they had to sort pictures and words as quickly as possible into two categories. Task order was counterbalanced so that half of the participants started with the IFT and vice versa. Participants were studied in a separate room to minimize external distractions. They were rewarded 5 euros in cash. Written informed consent was obtained pre-testing. Ethical approval was granted by the faculty ethical review board from the social and behavioral sciences faculty at Utrecht University, number 19–232.

Impression Formation Task

Stimuli

Twenty pictures of Caucasian children with a neutral facial expression (10 boys, 10 girls) were selected from the CAFE database (LoBue & Thrasher, 2015). The pictures selected have previously been used in a similar paradigm and represented the most clearly male-typed boys and female-typed girls (Endendijk et al.,

2019a). The pictures' brightness levels were adjusted so that the mean luminance ranged between 195 and 205.

Next to the face pictures, a stimulus set was created with 10 masculine toy words (crane, tractor, race car, garage, toolkit, soccer, digger, fire truck, pirates costume, helicopter) and 10 feminine toy words (tea set, princess dress, hula-hoop, doll clothing, barbie, play kitchen, jewelry, doll house). These toys were clearly defined as masculine and feminine in previous studies (Blakemore & Centers, 2005; Endendijk et al., 2014, 2019a). Thus far, studies on the neural correlates of implicit gender stereotypes have solely focused on stereotypes about boys' and girls' toy and activity preferences. However, adults also hold gendered expectations about young children's personality and behavior characteristics (Martin, 1995). Therefore, a second stimulus set was derived from the externalizing and internalizing behavior scales from the Child Behavior Checklist (CBCL; Achenbach, 1999). We selected 10 words reflecting internalizing behavior (dependent, shy, unhappy, depressed, sad, fearful, worried, ashamed, avoidant, sensitive) and 10 words reflecting externalizing behavior (violent, fighting, threatening, kicking, agitated, inattentive, noisy, cruel, disobedient, aggressive). To select these words, 55 young adults ($M_{age} = 22.2$, $SD = 2.52$, 43.6% male) were asked to rate 39 descriptions of behavior on a scale from 1 (highly typical for females) to 5 (highly typical for males). These ratings were analyzed with a one-sample T-test to see whether the descriptions were rated as significantly more male-typed or female-typed than the neutral mid-point of the scale (lowest M_{diff} : 0.255, $t(54) = -3.422$, $p = 0.001$, for unhappy/feminine; see Supplementary Table S1 for an overview of the CBCL words tested and T-statistics for each word). The selected words were rated as most male-typed or female-typed (i.e., largest mean difference from the neutral midpoint of the scale) by the adults. Male- and female-typed words were matched on number of syllables (2–4 syllables) in both the toy stimulus set and the behavior stimulus set.

Task design

The task consisted of two blocks, each containing 120 trials. In each trial, participants were presented with a picture of a child with a neutral facial expression, after which a toy (block 1) or behavior (block 2) word appeared. Participants were told in the first block that this word described a toy the child loves to play with and that they had to quickly form an impression of each child based on the information that was provided. In the second block, the instructions were that the word

described behavior that the child frequently exhibits. For half of the number of trials but randomly assigned, a question appeared after the face–word combination. This question was: How appropriate do you think this toy/behavior is for the child’s gender? Participants were instructed to evaluate the face–word combinations based on their first impressions. Answers were given by pressing 1–9 on the keyboard with higher scores indicating that the participant thought the toy/behavior was more appropriate.

The task took 18–22 minutes to complete, depending on the length of the self-paced break in between the two blocks. Each face picture appeared a total of 12 times, 6 times with a word that was congruent with gender stereotypes (e.g., boy face paired with the word “crane”, girl face paired with the word “sad”) and 6 times with gender-stereotype incongruent words (e.g., boy face paired with the word “tea set”, girl face paired with the word “aggressive”). The words were pseudo-randomly assigned to the child pictures, ensuring no word stimulus appeared twice with the same child face picture and each child face picture had appeared once, before being presented a second time. Each trial started with a fixation cross lasting for a varying amount of time (800, 900, 1000, 1100, or 1200 ms, randomly chosen) after which the face picture (1000 ms, width: 13.3 cm, height: 9.2 cm) appeared, superimposed on a gray background (191;191;191). After a randomly assigned jittered interstimulus interval (200, 225, 250, 275, or 300 ms) the word stimulus was presented in black (1000 ms, Cambria, font size 55). During half of the trials the question was presented (Cambria font size 24) and appeared until the participant had pressed a response-key (1–9). The task was presented electronically on a 14-inch laptop with the use of E-Prime v3.0 (Taylor & Marsh, 2017).

Explicit gender attitudes about toys or behavior

Appropriateness ratings during the IFT were extracted and used as a measure of the level of explicit gender attitudes about toys and behavior per participant. Explicit gender attitudes about toys were calculated by subtracting the average appropriateness score (1–9) on incongruent trials in the toy block from the average appropriateness score on congruent trials in the toy block. The same difference score was calculated for the behavior blocks. A high positive score meant that a participant evaluated stereotype-congruent child-toy combinations or child-behavior combinations as more appropriate than stereotype-incongruent stimulus combinations.

Implicit Association Test

A modified IAT that was previously used to measure violations of gender expectations about children’s toys in parents (Endendijk et al., 2013, 2019b) was extended by adding two blocks with behavior words. Participants were instructed to divide the stimuli, i.e., the toy pictures and behavior descriptions, between two children as quickly as possible, by means of pressing one of two keys (z, m) on the keyboard that were assigned to each child. Illustrations of the two children were presented continuously in the upper left and upper right corner of the monitor, superimposed on a white screen. The toy pictures and behavior descriptions were the same as described in the IFT.

The task consisted of four blocks (toy congruent; toy incongruent; behavior congruent; behavior incongruent) that each consisted of 68 trials and a practice block consisting of 20 trials. In the congruent blocks, participants were asked to assign feminine toys/behaviors to a girl and masculine toys/behaviors to a boy (e.g., assign a crane to a boy, assign “sad” to a girl). In the incongruent trials, the feminine toys/behaviors were assigned to a boy and the masculine toys/behaviors to a girl (e.g., assign a crane to a girl, assign “sad” to a boy). Participants were given a short description of the children in the beginning of each block. For instance, in the toy block the following instruction was given: “This is Linda. Linda loves dolls, barbies, princesses, playing with her hula-hoop and her play kitchen”. In the behavior block and example of the instruction was: “This is Kees. Kees is shy and dependent. He is easily embarrassed and very sensitive. He is often sad, fearful and depressed”. Clear exemplars of feminine and masculine toys and behaviors were chosen that covered the range of toy and behavior stimuli that would be used in that category. In the toy blocks, full-color toy pictures were presented in the middle of the screen until the participant pressed a response key. In the behavior blocks, behavior descriptions were presented in black (Courier New, font size 34).

Trials were separated by a 500 ms interstimulus interval. Jitter was created by the participants’ varying response times ($M = 902.62$, $SD = 695.62$, range: 18–9244 ms). The order of blocks was counterbalanced, so that half of the participants started with a congruent block, and the other half of the participants started with an incongruent block. Participants were given a self-paced break between each block when new instructions were given. The task took 12–15 minutes to complete.

Implicit gender stereotypes

The improved scoring algorithm for the IAT was used to calculate the level of implicit gender stereotypes of the participant (Greenwald et al., 2003). A high positive score represented more difficulties (e.g., longer reaction times, more errors) during incongruent trials compared to congruent trials, indicating more stereotyped expectations about the appropriateness of certain toys and behaviors for girls and boys.

EEG recordings

During both tasks, EEG was recorded from 32 scalp sites with the use of BioSemi ActiveTwo Ag-AgCl pin electrodes and hardware (BioSemi, 2011). The electrodes were placed according to the 10–20 electrode system with use of a nylon electrode cap (Klem et al., 1999). EEG signals were amplified, bandpass filtered at DC–400 Hz and sampled at 2048 Hz. Eye movements were recorded with four sintered Ag-AgCl electrodes placed above and below the right eye and just outside the outer corners of each eye.

EEG data of each task was processed separately, but because of the large similarity in processing we only indicate which steps differed between the tasks. Data were downsampled to a 256 Hz sampling rate, after which the data were bandpass filtered between 0.1 and 30 Hz (IFT) and 4 and 30 Hz (IAT). Individual participants' data were re-referenced to the average activity in all channels. Ocular artifacts were corrected using the Gratton and Coles method as implemented in Brainvision Analyzer (Gratton et al., 1983). Data were then segmented into epochs of –200 ms to 1000 ms, time-locked to the onset of the toy and behavior stimuli. For the IFT, a baseline correction was applied from –200 to 0 to correct for differences in absolute voltage and drift between trials and electrodes. This was not done for the IAT paradigm in accordance with previous processing in ERP studies (Endendijk et al., 2019b; Forbes et al., 2012). By baselining, between-subject differences in

preparatory activity would reduce variance in post-stimulus ERP amplitudes. However, this preparatory activity is characteristic to the IAT block design in which congruent and incongruent trials are separated. Therefore, as recommended in previous work, we decided to use high-pass filtering with 4 Hz instead of 0.1 Hz to decrease issues with systematic differences in pre-stimulus baseline activity (Bidet-Caulet et al., 2012; Endendijk et al., 2019b; Healy et al., 2015). Consequently, late-ERP components may be less reliable, since they mainly consist of lower EEG frequencies (Demiralp et al., 2001). Artifacts were rejected semi-automatically. Trials were marked as bad and manually inspected if the voltage step exceeded 50 $\mu\text{V}/\text{ms}$, with a maximum allowed difference of values in intervals of 100 μV within a 200 ms window, or with a lower activity in intervals of 0.5 μV . Trials were discarded if, within the marked trial, the artifact appeared across two or more electrodes or on the electrode at one of the sites of interest. A channel was marked “bad” for that participant if artifacts were present in more than 25% of the trials. Channels that were marked as “bad” were discarded from preprocessing and further analysis. Participants were excluded if they had significant noise in more than 25% of the trials on multiple channels (see the supplementary materials for more details on data cleaning). The remaining data for each individual participant was averaged into eight grand average waveforms per condition (toy boy congruent, toy boy incongruent, toy girl congruent, toy girl incongruent, behavior boy congruent, behavior boy incongruent, behavior girl congruent, behavior girl incongruent) for each task. Finally, total average waveforms per condition were created from the grand average waveforms per participant.

ERPs

Grand average waveforms were visually inspected to select time windows and electrodes for the ERP components of interest. This was done per task and per block, since the two blocks differed in the type of words presented. The electrodes and time windows

Table 1. Selected Time Windows and Electrodes per ERP per Block.

Task	ERP	Toy block		Behavior block		In accordance with
		time window	electrodes	time window	electrodes	
IFT	P1	95–140	PO3, PO4, O1, O2, Oz	90–130	PO3, PO4, O1, O2, Oz	Dickter and Gyurovski (2012); He et al. (2009).
	N1	165–180	PO3, PO4, O1, O2	165–180	PO3, PO4, O1, O2	Dickter and Gyurovski (2012).
	P2	185–205	Oz	185–205	Oz, O1, O2	Dickter and Gyurovski (2012); He et al. (2009).
	LPP	450–650	Pz, PO3, PO4	360–550	Fz, Cz, FC1, FC2	Ito et al. (2004); Williams and Themanson (2011).
IAT	N2	300–360	Fz, Cz, FC1, FC2	285–340	Fz, Cz, FC1, FC2	Endendijk et al. (2019b).
	P3	360–420	Fz, Cz, FC1, FC2	375–425	Fz, Cz, FC1, FC2	Endendijk et al. (2019b).

ERP = event-related potential. IFT = Impression Formation Task. IAT = Implicit Association Test.

with the largest amplitudes were selected; see, Table 1 for the overview and Figure S1-S4 for the grand average waveforms for the selected electrodes per task and block.

Statistical analyses

Average waveform amplitudes per subject were imported in RStudio v.1.4.17 (R Core Team, 2013). The data were analyzed using multilevel modeling with the *lme4* package (Bates et al., 2018, 2007) with maximum likelihood (ML). Since we expected participants to differ in their neural activation patterns irrespective of task and condition, we chose to allow intercepts and slopes to vary per participant in our models (i.e., participant ID was a random factor in the models). Additionally, we used multiple electrodes to quantify each ERP component, and activity can be expected to vary across electrodes. We therefore included electrode as an additional random factor in our models. The average amplitudes were outcome variables in the models, and congruence and gender of the child in the picture (stimulus gender) were the main variables of interest and thus added as predictors. Gender of the participant and gender stereotypes and attitudes were additionally added as possible covariates. The Akaike Information Criterion (AIC; Bozdogan, 1987) was used to determine model fits. Model fit was compared between models with the *anova* function in R, and the log likelihood ratio test was used to assess model significance. Separate analyses were performed for each task, within which ERPs for toy and behavior stimuli were individually modeled. The main factor of interest was congruence, as this factor provided information on the differential neural processing of stimuli that violated versus confirmed gender expectations. Interaction terms were also added between the factor of interest congruence and stimulus gender, participant gender, implicit gender stereotypes and explicit gender attitudes. This resulted in the following model entered in R:

$$\begin{aligned} & \text{Amplitude Congruence} + \text{Stimulus.Gender} \\ & + \text{Impl.stereotype} + \text{Part.Gender} \\ & + \text{Expl.attitudes} + \text{Congruence} * \text{Stimulus.Gender} \\ & + \text{Congruence} * \text{Impl.stereotype} \\ & + \text{Congruence} * \text{Part.Gender} \\ & + \text{Congruence} * \text{Expl.attitudes} + \\ & (\text{Congruence} * \text{Stimulus.Gender} | \text{PartID}) \\ & + (1 | \text{PartID} : \text{Electrode}) \end{aligned}$$

We used Cook's distance to detect outliers per component. Influential cases were defined as being more than the 95th percentile. Outliers were substituted by the value of the 95th percentile and analyses were repeated with and without outliers to see if they affected the results. Including them changed the slope, but not the significance of the results. Outliers were thus corrected and included. Interaction effects were post-hoc analyzed with the use of a paired sample t-test. Effect sizes were estimated by calculating Pearson correlations, with $r = 0.1$ representing a small effect size, $r = 0.3$ a medium effect size, and $r = 0.5$ a large effect size (Cohen, 2013).

Results

Descriptive statistics

The data were checked on linearity of predictors, homogeneity of variance, multicollinearity, normally distributed errors, and outliers. Two participants were excluded from the IAT analyses due to excessive noise in the EEG recordings. No participants were excluded from the IFT analyses. We additionally assessed whether men and women differed in terms of background variables; see, Table 2 for mean and standard deviations of the variables of interest. Men were on average older than women ($t(25) = 44.696$, $p < 0.001$). Men and women had comparably high educational levels, ($F(23) = 0.309$, $p = 0.584$) and did not significantly differ in their average levels of implicit gender stereotypes as measured with the IAT ($F(22) = 0.366$, $p = 0.551$). For both men and women, average appropriateness ratings were higher for congruent than

Table 2. Correlation Matrix Between Factors of Interest.

Correlation matrix					Men			Women		
	Age	IGS	GA toy	GA behav	<i>M</i>	<i>SD</i>	Range	<i>M</i>	<i>SD</i>	Range
Age	-				27.50	2.59	24–31	24.20	1.69	22–27
IGS	−0.128	-			1.06	0.59	−0.37–1.67	1.13	0.40	0.27–1.66
GA toy	0.017	0.338	-		2.75	1.39	0.33–4.93	1.66	1.38	−1.03–3.93
GA behav	−0.103	0.316	0.270	-	1.06	0.76	−0.33–2.63	1.03	0.42	0.37–1.73

IGS = implicit gender stereotypes, GA = gender attitudes, behav = behavior

Table 3. Overview of Findings per Task, per Block and for each ERP.

Task	Block	ERP	Congruence effect	Dependent on
IFT	Toy	P1	no	
		N1	no	
		P2	no	
		LPP	no	
	Behavior	P1	yes	Implicit stereotypes
		N1	yes	Gender stimuli
		P2	no	
		LPP	yes	Gender attitudes
IAT	Toy	N2	no	
		P3	no	
	Behavior	N2	no	
		P3	yes	Implicit stereotypes

ERP = event-related potential. IFT = Impression Formation Task.
IAT = Implicit Association Test, LPP = late positive potential.

incongruent trials in the toy block (men: $M_{congruent} = 6.65$, $SD_{congruent} = 1.28$, $M_{incongruent} = 3.90$, $SD_{incongruent} = 1.32$; women: $M_{congruent} = 6.83$, $SD_{congruent} = 0.58$, $M_{incongruent} = 5.17$, $SD_{incongruent} = 1.12$) and the behavior block (men: $M_{congruent} = 5.39$, $SD_{congruent} = 0.72$, $M_{incongruent} = 4.32$, $SD_{incongruent} = 0.96$; women: $M_{congruent} = 5.27$, $SD_{congruent} = 0.41$, $M_{incongruent} = 4.23$, $SD_{incongruent} = 0.49$). Men differed more in their evaluation of congruent versus incongruent trials than women, which indicated that men held more traditional gender attitudes about children's toys ($t(48) = 5.902$, $p < 0.001$) and behavior ($t(48) = 3.788$, $p < 0.001$) than women.

IFT

In Table 3 an overview of all findings is presented separately for the IFT and IAT, showing effects of congruence, and whether congruence additionally interacted with a variable of interest on the ERP amplitudes.

Toy

P1

P1 amplitude was negatively associated with implicit gender stereotypes independent of stimulus type or congruence ($\beta = -0.315$, $t(25) = -2.177$, $p = 0.039$, $r = 0.399$). All other main effects and interactions with congruence were non-significant ($ps > 0.05$).

N1

Since there were some indications that P1 amplitude differences continued in N1 amplitudes, we corrected N1 amplitude for preceding P1 amplitudes in overlapping electrodes. The uncorrected N1 results can be found in the supplementary materials. A significant main effect for explicit gender attitudes was found ($\beta = 0.250$, $t(25) = 2.072$, $p = 0.049$, $r = 0.383$). Stronger

gendered attitudes were associated with larger N1 amplitudes regardless of congruence or stimulus gender. All other main effects and interactions were non-significant ($ps > 0.05$).

P2

P2 amplitude was negatively associated with implicit gender stereotypes ($\beta = -0.448$, $t(25) = -2.823$, $p = 0.009$, $r = 0.492$), irrespective of stimulus type or congruence. All other effects were non-significant ($ps > 0.05$).

LPP

LPP amplitudes were negatively related to implicit gender stereotypes irrespective of stimulus type or congruence ($\beta = -0.424$, $t(25) = -3.171$, $p = 0.004$, $r = 0.536$). All other effects were non-significant ($ps > 0.05$).

Behavior

P1

A significant main effect was found of implicit gender stereotypes ($\beta = -0.318$, $t(25) = -2.110$, $p = 0.045$, $r = 0.389$), as well as a significant interaction between congruence and implicit stereotypes ($\beta = 0.165$, $t(25) = 3.064$, $p = 0.005$, $r = 0.522$). Inspecting the data revealed that less strong gender stereotypes were associated with a larger P1 to congruent versus incongruent trials, whereas stronger implicit gender stereotypes were associated with a larger P1 toward incongruent versus congruent trials (see, Figure 1). All other effects and interactions were non-significant (p -values > 0.05).

N1. Since there were some indications that P1 amplitude effects carried over to N1 amplitudes, P1 amplitudes were added as a covariate to the analysis of N1 amplitudes. A significant interaction was found between stimulus type and congruence ($\beta = 0.166$, $t(25) = 2.467$, $p = 0.021$, $r = 0.442$). Post-hoc analyses revealed that N1 amplitude was significantly larger for boys' pictures paired with incongruent behavior (i.e., internalizing) than when paired with congruent behavior (i.e., externalizing) ($t(99) = 4.433$, $p < 0.001$, $r = 0.407$) but this difference was not found for girls' pictures ($t(99) = -1.558$, $p = 0.123$). In addition, congruent-girl trials elicited a significantly larger N1 effect than congruent-boy trials ($t(99) = 3.333$, $p = 0.001$, $r = 0.318$), whereas incongruent-boy trials elicited a significantly larger N1 effect than incongruent-girl trials ($t(99) = -3.092$, $p = 0.003$, $r = 0.297$; see, Figure 2).

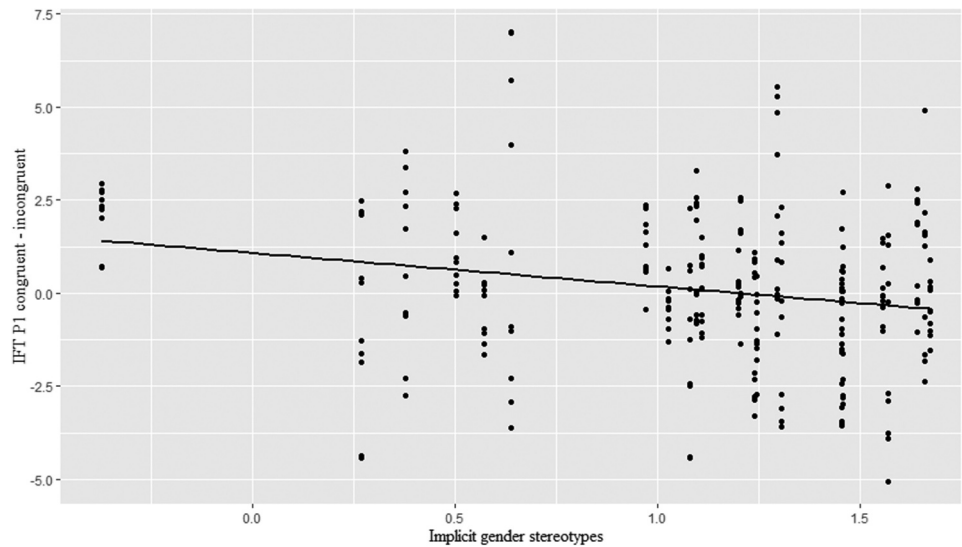


Figure 1. The Effect of Implicit Gender Stereotypes on the Difference Scores of P1 Amplitudes During the Behavior Block of the Impression Formation Task, as Calculated by Subtracting Incongruent Trials from Congruent Trials.

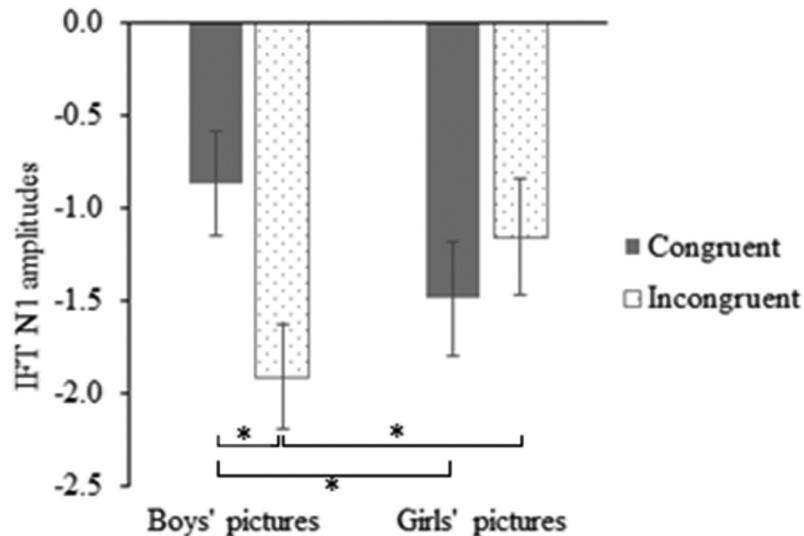


Figure 2. N1 Amplitude During an IFT to Stimuli That Violated Gendered Expectations About Behavior (i.e., incongruent) and Stimuli That Confirmed Gendered Expectations About Behavior (i.e., congruent), Separate for Boys' Pictures and Girls' Pictures. Note: The asterisk indicates a statistically significant difference between trials.

P2

The P2 model resulted in a significant negative main effect of implicit gender stereotypes on P2 amplitudes ($\beta = -0.369$, $t(25) = -2.279$, $p = 0.032$, $r = 0.415$). All other effects were non-significant (smallest p -value = 0.053 for a main effect of congruence).

LPP

A significant main effect was found for explicit gender attitudes ($\beta = 0.221$, $t(25) = 2.150$, $p = 0.042$, $r = 0.395$). Second, a significant interaction was

found between congruence and explicit gender attitudes ($\beta = -0.098$, $t(25) = -2.115$, $p = 0.045$, $r = 0.390$). As depicted in Figure 3, more traditional gender attitudes about behavior were associated with larger LPP amplitudes in response to congruent compared to incongruent trials, whereas less traditional gender attitudes were associated with larger LPP amplitudes in response to incongruent compared to congruent trials. All other effects and interactions were non-significant ($ps > 0.05$).

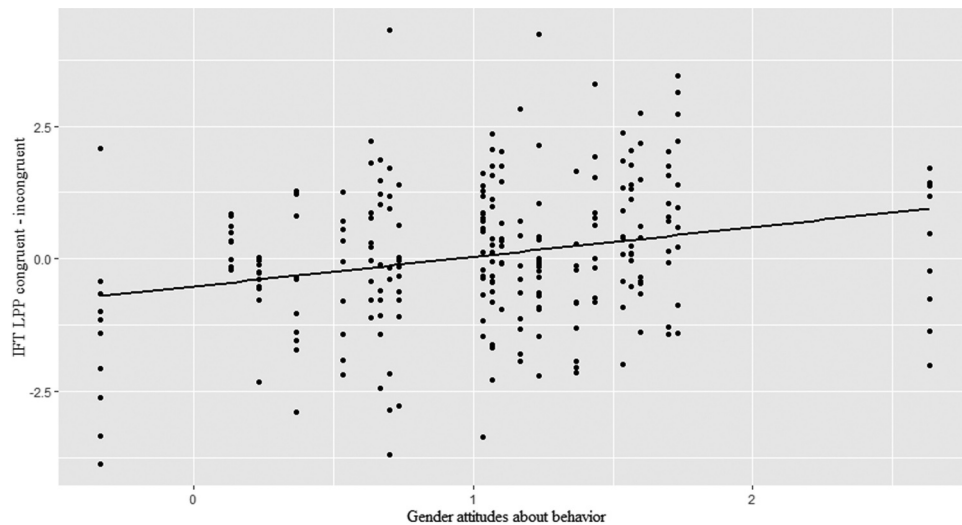


Figure 3. The Effect of Gender Attitudes About Behavior on the Difference Score of LPP Amplitudes During the Behavior Block of the Impression Formation Task, as Calculated by Subtracting Congruent from Incongruent Trials.

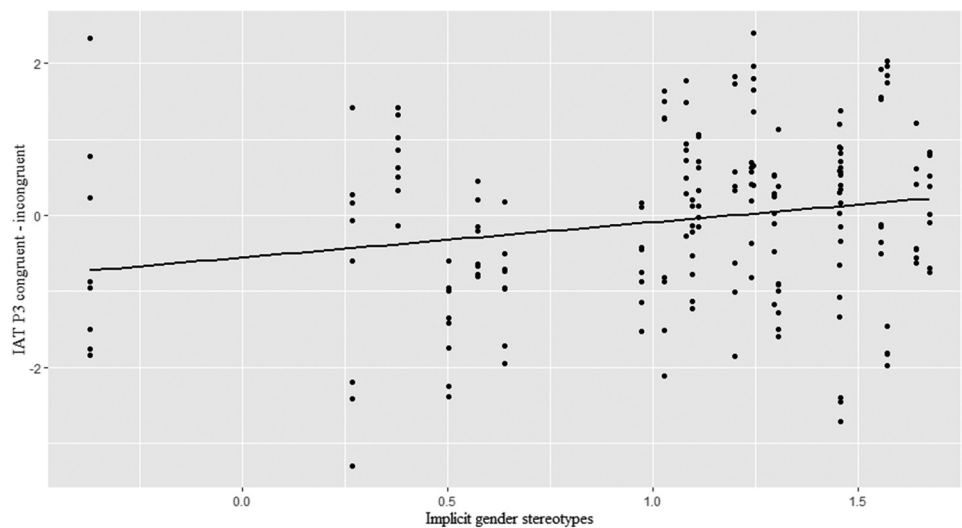


Figure 4. The Effect of Implicit Gender Stereotypes on the Difference Score of P3 Amplitudes During the Congruent vs. Incongruent Behavior Block of the Implicit Association Test, as Calculated by Subtracting Congruent from Incongruent Trials.

IAT

Toy

N2. No significant effects emerged from the N2 model (all p -values > 0.05).

P3. With regard to the P3, a significant main effect of stimulus type ($\beta = 0.322$, $t(23) = 2.422$, $p = 0.024$, $r = 0.451$) was found, indicating that assigning toys to

girls elicited a larger P3 amplitude than assigning toys to boys, regardless of which type of toy was presented.

Behavior

N2

The model assessing N2 amplitude effects did not yield any significant results (smallest p -value = 0.280 for main effect congruence).

P3

A significant interaction emerged between congruence and implicit stereotypes ($\beta = -0.191$, $t(23) = -2.603$, $p = 0.016$, $r = 0.477$; see, Figure 4). More traditional implicit gender stereotypes were associated with larger P3 amplitudes toward congruent than incongruent trials, whereas less traditional gender stereotypes were associated with larger P3 amplitudes toward incongruent than congruent trials.

Discussion

The current study examined the robustness of the neural activation patterns of two experimental paradigms when participants were either actively (IAT) or passively (IFT) exposed to visual stimuli that confirmed versus violated gender expectations. Although both tasks elicited significant ERP differences between stimuli that confirmed or violated gender expectations, the IFT evoked effects of stimulus' congruence with gender expectations on all but one ERP components of interest (P1, N1, and LPP). The IAT only showed an effect of stimulus congruence on P3 amplitude, and not on N2 amplitude. However, it must be noted that for the IFT, more ERP components were selected and tested than for the IAT. To evaluate whether the congruence effects we found with the IFT and IAT are relevant and interpretable in the context of expectancy violations, we will discuss them in light of previous research on the processes related to each ERP component.

Congruence ERPs elicited in the IFT

Regarding the IFT, a larger P1 to expectancy-violating vs. expectancy-confirming child-behavior stimuli was associated with stronger implicit gender stereotypes, whereas a larger P1 to expectancy-confirming vs. expectancy-violating stimuli was associated with less strong gender stereotypes. This finding indicated that stimuli that violate people's own stereotyped expectations about gender appear to be associated with increased visual processing (Luck, 2014). Previous research also demonstrated that stimuli that violated gender expectations elicited higher P1 amplitudes (Liu et al., 2017), but individual differences in gender stereotypes were not taken into account.

Next to the effects on P1, the IFT evoked effects of stimulus congruence on the N1 for the behavior trials. As previous research found larger N1 amplitudes for expectancy-violating stimuli compared to expectancy-confirming stimuli (Dickter & Gyurovski, 2012), our findings might indicate that boys paired with feminine behaviors violate gender expectations to a larger extent than

girls paired with masculine behavior. This fits with previous findings that gender stereotypes are more restrictive for boys than girls (Kane, 2006; Sandnabba & Ahlberg, 1999). We did not find differences in N1 effects during the toy block. This difference between the toy and behavior block might be attributed to the more negative connotation of the behavior stimuli compared to the toy stimuli. The N1 has been specifically implicated in research on negative words, with larger peaks reflecting early detection of relevant emotional information (Bernat et al., 2001).

Finally, the IFT elicited effects of stimulus congruence on the LPP for the behavior conditions of the task only. Larger LPP amplitudes in response to expectancy-confirming children were associated with participants' more traditional explicit gender attitudes (i.e., evaluating expectancy-confirming children as more appropriate than expectancy-violating children). Williams and Thernanson (2011) also found a larger LPP in expectancy-confirming trials of a gay-straight IAT. Larger LPP might reflect (controlled) attentional orienting to salient stimuli (e.g., Hajcak et al., 2013; Spreckelmeyer et al., 2006), which in the current study are children that show behavior that people evaluate as appropriate.

Congruence ERPs elicited in the IAT

By studying the activation patterns of the IAT, only the P3 was affected by congruence during the behavior block. However, the finding that P3 was larger for expectancy-confirming than expectancy-violating trials for people with stronger implicit gender stereotypes, was not in line with P3 being associated with attentional focus on unexpected items (Polich, 2007) or negatively valenced stimuli (Duval et al., 2013; Gyurovski et al., 2018). For people with strong implicit stereotypes expectancy-violating stimuli would constitute unexpected or negatively valenced stimuli, which we expected to elicit larger P3, similar to a previous study (Bartholow & Dickter, 2007). The unexpected findings for the P3 with the IAT might be due to the high-pass filtering that was necessary for the IAT data, which made late-ERPs (e.g., P3, N400) less reliable as these ERPs often consist of lower EEG frequencies in the delta band (0–4 Hz; Demiralp et al., 2001).

Other important findings

It is important to note that no main effect of congruence was found on ERP amplitudes, only interactions with other variables. Previous studies have indicated that ERP amplitude differences toward expectancy-violating and expectancy-confirming trials was dependent on the level

of implicit (gender) stereotypes (e.g., Endendijk et al., 2019b; Healy et al., 2015), hostile sexism (Canal et al., 2015) or participants' gender (Proverbio et al., 2018). When such interactions are found, between-group variation in ERPs toward congruent and incongruent trials may cancel each other out. The current results confirm the need to consider participants' level of gender stereotypes or gender attitudes in research on the neural processing of violations of gender expectations.

Another important finding of this study is that people's brains also seem to respond differently toward the violation of gendered expectations about behaviors versus toys. For instance, more consistent differences in brain responses were observed when people's gendered expectations about behaviors were violated vs. confirmed, than when people's expectations about toys were violated. People may have responded more strongly toward the violation of expectations about problem behaviors, because these behaviors have a negative connotation and bad events generally have a larger impact on people than good events (Baumeister et al., 2001). In addition, the findings regarding violations of gendered expectations about problem behavior extend previous research that focused on toys (Endendijk et al., 2019b), emotion expression (Liu et al., 2017), or gender-typed traits and activities (Proverbio et al., 2018; White et al., 2009). Adults generally expect boys to possess more masculine traits, such as being dominant, independent, competitive, and aggressive, whereas girls are expected to possess more feminine traits, such as being gentle, neat, sympathetic, weak, shy, overly emotional, and feminine looking (Koenig, 2018; Martin, 1995; Powlisha, 2000). Our results confirm that people hold gendered expectations about internalizing and externalizing behaviors, and that violation of these expectations is visible in distinct brain activity patterns. Future research could examine whether brain responses to the violation of gendered expectations about child problem behaviors may underlie why parents socialize boys and girls to differently express their emotions (i.e., internalizing or externalizing emotions; Chaplin et al., 2005).

Limitations and directions for future research

The findings from this study must be interpreted in light of its limitations. Firstly, the small sample size reduced the power for between-subjects comparisons. Thus, the interactions gender stereotypes and attitudes need to be interpreted with caution. Also, adding random slopes to the multilevel models greatly reduced the power to detect medium and small effects in congruence. Although random slopes are recommended to reduce the probability of type-I

errors (Heisig & Schaeffer, 2019), more power is needed to detect smaller effect sizes (Clayson et al., 2019). Future studies with larger sample sizes are therefore needed to confirm whether neural processing of gendered stimuli differs between male and female participants and varies with the strength of people's gender stereotypes and attitudes. In addition, different ERP components have been selected for the IFT and IAT, which hampers direct comparison between the two tasks. Moreover, we examined neural reactions to children violating gender expectations in non-parental adults. It could be that parents have stronger, or less strong, expectations when it comes to children, and thus may respond differently than non-parents when these expectations are violated. In this study, we have only examined the violation of social expectations with regard to (child) gender. Therefore, further research is necessary to confirm whether the IFT is more suitable than the IAT to assess violations of social expectations across other social domains than gender (e.g., race, age, sexual orientation). Nonetheless, we have been able to find more robust effects with the IFT, which encourages future research to examine the neural correlates of violations to social (gender) expectations with the use of an IFT.

Conclusion

The findings from the IFT show that this experimental paradigm is more suitable than the IAT to combine with EEG research on the neural processes underlying people's responses to violations of gender expectations. The IFT elicited more relevant ERP patterns in response to stimuli that violated social expectations than the IAT, with the latter experiencing some drawbacks in EEG processing due to its block design. Based on these findings, we recommend future research examining the neural processes underlying violations to social expectations to use impression formation paradigms, in which people are exposed to stimuli that violate and confirm social expectations.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

ORCID

Christel M. Portengen  <http://orcid.org/0000-0003-2437-1511>
 Anneloes L. van Baar  <http://orcid.org/0000-0002-3498-9019>
 Joyce J. Endendijk  <http://orcid.org/0000-0001-8149-912X>

References

- Achenbach, T. M. (1999). The child behavior checklist and related instruments. In M. E. Maruish (Ed.), *The use of psychological testing for treatment planning and outcome assessment* (pp. 429–466). Erlbaum.
- Azizian, A., Freitas, A. L., Parvaz, M. A., & Squires, N. K. (2006). Beware misleading cues: Perceptual similarity modulates the N2/P3 complex. *Psychophysiology*, 43(3), 253–260. <https://doi.org/10.1111/j.1469-8986.2006.00409.x>
- Bartholow, B. D., & Dickter, C. L. (2007). Social cognitive neuroscience of person perception: A selective review focused on the event-related brain potential. In E. Harmon-Jones & P. Winkielman (Eds.), *Social neuroscience: Integrating biological and psychological explanations of social behavior* (pp. 376–400). Guilford Press.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., Dai, B., Scheipl, F., Grothendieck, G., & Green, P. (2018). Package 'lme4'. Version, 1, 17.
- Bates, D., Sarkar, D., Bates, M. D., & Matrix, L. (2007). The lme4 package. *R package version*, 2(1), 74.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Bernat, E., Bunce, S., & Shevrin, H. (2001). Event-related brain potentials differentiate positive and negative mood adjectives during both supraliminal and subliminal visual processing. *International Journal of Psychophysiology*, 42(1), 11–34. [https://doi.org/10.1016/S0167-8760\(01\)00133-7](https://doi.org/10.1016/S0167-8760(01)00133-7)
- Bidet-Caulet, A., Barbe, P. G., Roux, S., Viswanath, H., Barthélémy, C., Bruneau, N., Knight, R. T., & Bonnet-Brilhault, F. (2012). Dynamics of anticipatory mechanisms during predictive context processing. *European Journal of Neuroscience*, 36(7), 2996–3004. <https://doi.org/10.1111/j.1460-9568.2012.08223.x>
- BioSemi. (2011). Biosemi activetwo.[eeg system]. [Mobile application software]
- Blakemore, J. E. O., Berenbaum, S. A., & Liben, L. S. (2008). *Gender development*. Psychology Press.
- Blakemore, J. E. O., & Centers, R. E. (2005). Characteristics of boys' and girls' toys. *Sex Roles*, 53(9–10), 619–633. <https://doi.org/10.1007/s11199-005-7729-0>
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika*, 52(3), 345–370. <https://doi.org/10.1007/BF02294361>
- Canal, P., Garnham, A., & Oakhill, J. (2015). Beyond gender stereotypes in language comprehension: Self sex-role descriptions affect the brain's potentials associated with agreement processing. *Frontiers in Psychology*, 6, 1953. <https://doi.org/10.3389/fpsyg.2015.01953>
- Chaplin, T. M., & Aldao, A. (2013). Gender differences in emotion expression in children: A meta-analytic review. *Psychological Bulletin*, 139(4), 735–765. <https://doi.org/10.1037/a0030737>
- Chaplin, T. M., Cole, P. M., & Zahn-Waxler, C. (2005). Parental socialization of emotion expression: Gender differences and relations to child adjustment. *Emotion*, 5(1), 80–88. <https://doi.org/10.1037/1528-3542.5.1.80>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic press.
- P. E. Carbine, K. A., Baldwin, S. A., & Larson, M. J. (2019). Methodological reporting behavior, sample sizes, and statistical power in studies of event-related potentials: Barriers to reproducibility and replicability. *Psychophysiology*, 56(11), e13437. [10.1111/psyp.13437](https://doi.org/10.1111/psyp.13437)
- Demiralp, T., Ademoglu, A., Comerchero, M., & Polich, J. (2001). Wavelet analysis of P3a and P3b. *Brain Topography*, 13(4), 251–267. <https://doi.org/10.1023/A:1011102628306>
- Dickter, C., & Gyurovski, I. (2012). The effects of expectancy violations on early attention to race in an impression-formation paradigm. *Social Neuroscience*, 7(3), 240–251. <https://doi.org/10.1080/17470919.2011.609906>
- Duval, E. R., Moser, J. S., Huppert, J. D., & Simons, R. F. (2013). What's in a face? *Journal of Psychophysiology*, 27(1), 27–38. <https://doi.org/10.1027/0269-8803/a000083>
- 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), 577–590. <https://doi.org/10.1007/s11199-013-0265-4>
- Endendijk, J. J., Groeneveld, M. G., Van der Pol, L. D., Van Berkel, S. R., Hallers-Haalboom, E. T., Mesman, J., & Bakermans-Kranenburg, M. J. (2014). Boys don't play with dolls: Mothers' and fathers' gender talk during picture book reading. *Parenting*, 14(3–4), 141–161. <https://doi.org/10.1080/15295192.2014.972753>
- Endendijk, J. J., Smit, A. K., Van Baar, A. L., & Bos, P. A. (2019a). Boys' toys, girls' toys: An fMRI study of mothers' neural responses to children violating gender expectations. *Biological Psychology*, 148, 107776. <https://doi.org/10.1016/j.biopsycho.2019.107776>
- Endendijk, J. J., Spencer, H., Bos, P. A., & Derks, B. (2019b). Neural processing of gendered information is more robustly associated with mothers' gendered communication with children than mothers' implicit and explicit gender stereotypes. *Social Neuroscience*, 14(3), 300–312. <https://doi.org/10.1080/17470919.2018.1468357>
- Fabre, E. F., Causse, M., Pesciarelli, F., & Cacciari, C. (2015). Sex and the money – How gender stereotypes modulate economic decision-making: An ERP study. *Neuropsychologia*, 75, 221–232. <https://doi.org/10.1016/j.neuropsychologia.2015.06.013>
- Forbes, C. E., Cameron, K. A., Grafman, J., Barbey, A. K., Solomon, J., Ritter, W., & Ruchkin, D. (2012). Identifying temporal and causal contributions of neural processes underlying the Implicit Association Test (IAT). *Frontiers in Human Neuroscience*, 6, 320. <https://doi.org/10.3389/fnhum.2012.00320>
- Friedman, C. K., Leaper, C., & Bigler, R. S. (2007). Do mothers' gender-related attitudes or comments predict young children's gender beliefs? *Parenting: Science & Practice*, 7(4), 357–366. <https://doi.org/10.1080/15295190701665656>

- Gratton, G., Coles, M. G., & Donchin, E. (1983). A new method for off-line removal of ocular artifact. *Electroencephalography & Clinical Neurophysiology*, 55(4), 468–484. [https://doi.org/10.1016/0013-4694\(83\)90135-9](https://doi.org/10.1016/0013-4694(83)90135-9)
- Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, 85(2), 197. <https://doi.org/10.1037/0022-3514.85.2.197>
- 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. <https://doi.org/10.1037/a0015575>
- Gyurovski, I., Kubota, J., Cardenas-Iniguez, C., & Cloutier, J. (2018). Social status level and dimension interactively influence person evaluations indexed by P300s. *Social Neuroscience*, 13(3), 333–345. <https://doi.org/10.1080/17470919.2017.1326400>
- Haines, E. L., Deaux, K., & Lofaro, N. (2016). The times are a-changing ... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3), 353–363. <https://doi.org/10.1177/0361684316634081>
- Hajcak, G., MacNamara, A., Foti, D., Ferri, J., & Keil, A. (2013). The dynamic allocation of attention to emotion: Simultaneous and independent evidence from the late positive potential and steady state visual evoked potentials. *Biological Psychology*, 92(3), 447–455. <https://doi.org/10.1016/j.biopsycho.2011.11.012>
- He, Y., Johnson, M. K., Dovidio, J. F., & McCarthy, G. (2009). The relation between race-related implicit associations and scalp-recorded neural activity evoked by faces from different races. *Social Neuroscience*, 4(5), 426–442. <https://doi.org/10.1080/17470910902949184>
- Healy, G. F., Boran, L., & Smeaton, A. F. (2015). Neural patterns of the Implicit Association Test. *Frontiers in Human Neuroscience*, 9, 605. <https://doi.org/10.3389/fnhum.2015.00605>
- Helman, E., Ingbreten, Z. A., & Freeman, J. B. (2014). The neural basis of stereotypic impact on multiple social categorization. *NeuroImage*, 101, 704–711. <https://doi.org/10.1016/j.neuroimage.2014.07.056>
- Heilman, M. E., Wallen, A. S., Fuchs, D., & Tamkins, M. M. (2004). Penalties for success: Reactions to women who succeed at male gender-typed tasks. *Journal of Applied Psychology*, 89(3), 416. <https://doi.org/10.1037/0021-9010.89.3.416>
- Heisig, J. P., & Schaeffer, M. (2019). Why you should always include a random slope for the lower-level variable involved in a cross-level interaction. *European Sociological Review*, 35(2), 258–279. <https://doi.org/10.1093/esr/jcy053>
- Hilgard, J., Bartholow, B. D., Dickter, C. L., & Blanton, H. (2014). Characterizing switching and congruency effects in the Implicit Association Test as reactive and proactive cognitive control. *Social Cognitive and Affective Neuroscience*, 10(3), 381–388. <https://doi.org/10.1093/scan/nsu060>
- Ito, T. A., Thompson, E., & Cacioppo, J. T. (2004). Tracking the timecourse of social perception: The effects of racial cues on event-related brain potentials. *Personality and Social Psychology Bulletin*, 30(10), 1267–1280. <https://doi.org/10.1177/0146167204264335>
- Jerónimo, R., Volpert, H. I., & Bartholow, B. D. (2017). Event-related potentials reveal early attention bias for negative, unexpected behavior. *Social Neuroscience*, 12(2), 232–236. <https://doi.org/10.1080/17470919.2016.1144646>
- Kane, E. W. (2006). “No way my boys are going to be like that!” Parents’ responses to children’s gender nonconformity. *Gender & Society*, 20(2), 149–176. <https://doi.org/10.1177/0891243205284276>
- Klem, G., Luders, H. O., Jasperf, H., & Elger, C. (1999). Section 1. EEG (Technical Standards and Glossary)-1.1. The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology-Supplements Only*, 52, 3–6.
- Koenig, A. M. (2018). Comparing prescriptive and descriptive gender stereotypes about children, adults, and the elderly. *Frontiers in Psychology*, 9, 1086. <https://doi.org/10.3389/fpsyg.2018.01086>
- Li, T., Cardenas-Iniguez, C., Correll, J., & Cloutier, J. (2016). The impact of motivation on race-based impression formation. *NeuroImage*, 124, 1–7. <https://doi.org/10.1016/j.neuroimage.2015.08.035>
- Liu, C., Liu, Y., Iqbal, Z., Li, W., Lv, B., & Jiang, Z. (2017). Symmetrical and asymmetrical interactions between facial expressions and gender information in face perception. *Frontiers in Psychology*, 8, 1383. <https://doi.org/10.3389/fpsyg.2017.01383>
- LoBue, V., & Thrasher, C. (2015). The child affective facial expression (CAFE) set: Validity and reliability from untrained adults. *Frontiers in Psychology*, 5, 1532. <https://doi.org/10.3389/fpsyg.2014.01532>
- Luck, S. J. (2014). *An introduction to the event-related potential technique*. MIT press.
- Luck, S. J., Woodman, G. F., & Vogel, E. K. (2000). Event-related potential studies of attention. *Trends in Cognitive Sciences*, 4(11), 432–440. [https://doi.org/10.1016/S1364-6613\(00\)01545-X](https://doi.org/10.1016/S1364-6613(00)01545-X)
- Maas, C. J., & Hox, J. J. (2004). The influence of violations of assumptions on multilevel parameter estimates and their standard errors. *Computational Statistics & Data Analysis*, 46(3), 427–440. <https://doi.org/10.1016/j.csda.2003.08.006>
- Martin, C. L. (1995). Stereotypes about children with traditional and nontraditional gender roles. *Sex Roles*, 33(11–12), 727–751. <https://doi.org/10.1007/bf01544776>
- Osterhout, L., Bersick, M., & McLaughlin, J. (1997). Brain potentials reflect violations of gender stereotypes. *Memory & Cognition*, 25(3), 273–285. <https://doi.org/10.3758/BF03211283>
- Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. *Clinical Neurophysiology*, 118(10), 2128–2148. <https://doi.org/10.1016/j.clinph.2007.04.019>
- Powlishta, K. K. (2000). The effect of target age on the activation of gender stereotypes. *Sex Roles*, 42(3), 271–282. <https://doi.org/10.1023/A:1007095307977>
- Proverbio, A. M., Alberio, A., & De Benedetto, F. (2018). Neural correlates of automatic beliefs about gender stereotypes: Males are more prejudicial. *Brain & Language*, 186, 8–16. <https://doi.org/10.1016/j.bandl.2018.08.006>
- Proverbio, A. M., Orlandi, A., & Bianchi, E. (2017). Electrophysiological markers of prejudice related to sexual gender. *Neuroscience*, 358, 1–12. <https://doi.org/10.1016/j.neuroscience.2017.06.028>
- R Core Team. (2013). *R: A language and environment for statistical computing*.

- Rodríguez-Gómez, P., Romero-Ferreiro, V., Pozo, M. A., Hinojosa, J. A., & Moreno, E. M. (2020). Facing stereotypes: ERP responses to male and female faces after gender-stereotyped statements. *Social Cognitive and Affective Neuroscience*, 15(9), 928–940. <https://doi.org/10.1093/scan/nsaa117>
- Sandnabba, N. K., & Ahlberg, C. (1999). Parents' attitudes and expectations about children's cross-gender behavior. *Sex Roles*, 40(3–4), 249–263. <https://doi.org/10.1023/A:1018851005631>
- Schindler, S., Bruchmann, M., & Straube, T. (2020). Imagined veridicality of social feedback amplifies early and late brain responses. *Social Neuroscience*, 15(6), 678–687. <https://doi.org/10.1080/17470919.2020.1857303>
- Siyanova-Chanturia, A., Pesciarelli, F., & Cacciari, C. (2012). The electrophysiological underpinnings of processing gender stereotypes in language. *PloS One*, 7(12), e48712. <https://doi.org/10.1371/journal.pone.0048712>
- Spreckelmeyer, K. N., Kutas, M., Urbach, T. P., Altenmüller, E., & Münte, T. F. (2006). Combined perception of emotion in pictures and musical sounds. *Brain Research*, 1070(1), 160–170. <https://doi.org/10.1016/j.brainres.2005.11.075>
- Sullivan, J., Moss-Racusin, C., Lopez, M., & Williams, K. (2018). Backlash against gender stereotype-violating preschool children. *PloS One*, 13(4), e0195503. <https://doi.org/10.1371/journal.pone.0195503>
- Taylor, M. J. (2002). Non-spatial attentional effects on P1. *Clinical Neurophysiology*, 113(12), 1903–1908. [https://doi.org/10.1016/S1388-2457\(02\)00309-7](https://doi.org/10.1016/S1388-2457(02)00309-7)
- Taylor, P. J., & Marsh, J. E. (2017). E-Prime (Software). In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods*. <https://doi.org/10.1002/9781118901731.iecrm0085>
- Vogel, E. K., & Luck, S. J. (2000). The visual N1 component as an index of a discrimination process. *Psychophysiology*, 37(2), 190–203. <https://doi.org/10.1111/1469-8986.3720190>
- White, K. R., Crites, S. L., Jr, Taylor, J. H., & Corral, G. (2009). Wait, what? Assessing stereotype incongruities using the N400 ERP component. *Social Cognitive and Affective Neuroscience*, 4(2), 191–198. <https://doi.org/10.1093/scan/nsp004>
- Williams, J. K., & Themanson, J. R. (2011). Neural correlates of the Implicit Association Test: Evidence for semantic and emotional processing. *Social Cognitive and Affective Neuroscience*, 6(4), 468–476. <https://doi.org/10.1093/scan/nsq065>
- Xiao, F., Zheng, Z., Wang, Y., Cui, J., & Chen, Y. (2015). Conflict monitoring and stimulus categorization processes involved in the prosocial attitude Implicit Association Test: Evidence from event-related potentials. *Social Neuroscience*, 10(4), 408–417. <https://doi.org/10.1080/17470919.2014.1003598>
- Yang, Y., White, K. R. G., Fan, X., Xu, Q., & Chen, Q.-W. (2020). Differences in explicit stereotype activation among social groups based on the stereotype content model: Behavioral and electrophysiological evidence in Chinese sample. *Brain Sciences*, 10(12), 1001. <https://doi.org/10.3390/brainsci10121001>
- Yao, Z., & Wang, Z. (2014). Concreteness of positive word contributions to affective priming: An ERP study. *International Journal of Psychophysiology*, 93(3), 275–282. <https://doi.org/10.1016/j.ijpsycho.2014.06.005>References