
Usage Guidelines:
Please refer to usage guidelines at lib-eprints@bbk.ac.uk. or alternatively contact lib-eprints@bbk.ac.uk.
S1. Supplementary Methods

S1.1. Participant characteristics

Inclusion criteria for high-risk infant siblings included age (≤ 6 months), presence of autism in a full biological older sibling, and anticipated residence in the region (within 1.5 hours driving distance from the University) for the next 2 years. To confirm the diagnosis of ASD in an older sibling, the Autism Diagnostic Interview-Revised (ADI-R) was administered by phone and medical records were collected to confirm the diagnosis was based on DSM-IV criteria from a psychologist or physician.

Inclusion criteria for the low risk longitudinal group (Controls_long) included age (≤ 6 months) and a biological older sibling without a diagnosis of ASD. Inclusion criteria for low-risk cross sectional infants (Controls_cross) included age (6 or 12 months); infants were not required to have an older sibling. For the Controls_cross more infants were enrolled in the EEG and ERP tasks (6 months: n=114, 51 females, 12 months: n=104, 50 female) that the behavioral attention tasks (6 months: n=51, 27 females, 12 months: n=54, 27 female) given the higher attrition rate for EEG in this age range. Exclusionary criteria for both low-risk groups included a known family history of ASD in 1st or 2nd degree relatives.

Additional exclusion criteria for the all participant groups included: physical signs (e.g., dysmorphic features) of known genetic syndromes, serious medical or neurological conditions (e.g., encephalitis, concussion, seizure disorder, diabetes, congenital heart disease), neurocutaneous markings, or sensory impairments such as vision or hearing loss; serious motor impairment; birth weight < 2000 grams and/or gestational age < 37 wks, history of intraventricular hemorrhage, exposure to neurotoxins (including alcohol, drugs), and maternal gestational diabetes. In addition,
variables that may impact family functioning (e.g., serious parental substance abuse, bipolar disorder, or psychosis) were exclusion criteria.

**S1.2. Further details of the Promoting First Relationships intervention.**

*Previous evidence of Efficacy*

PFR has been evaluated in two randomized clinical trials: the Fostering Families Project [FFP] (MH077329, Susan Spieker- PI) and Supporting Parents Program [SPP] (HD0610362, Monica Oxford- PI) in two child welfare samples in Washington State. Both studies trained community providers to deliver the 10-week program, served children 10 to 24 months of age, and were focused on children in the child welfare system.

The **FFP** project was for toddlers who had been recently separated from a caregiver (either being removed from the birth home to foster care or kin care, or being returned to the birth home after a stay in foster or kin care):

- **Findings:** Caregivers who received PFR improved in their sensitive parenting while the children showed increased competence (Spieker et al., 2012) improved cortisol functioning (Nelson & Spieker, 2013). Children who were with foster care or kin providers were more likely to be adopted (Spieker et al., 2014). Children, in the PFR condition, who returned to their birth-home, showed improved sleep, this effect was mediated by a decrease in separation anxiety (Oxford et al., 2013); reunified birth parents also showed improvement in observed parenting sensitivity and child regulation as well as parent self-report parenting and child behavior (Oxford, Marcenko, et al., in press).
The SPP study was aimed at those families who were being investigated for child maltreatment in Snohomish and Skagit counties in Washington State (Oxford et al., 2015). Eligibility was to have an open case in CPS. Preliminary results are in the process of being published but include:

- A significant reduction in foster care placements for those receiving PFR (6%) compared to those in the resource and referral group (13%).
- A significant improvement in observed parental sensitivity and parental child development knowledge.
- A significant reduction in observed child based atypical affective communications.

Reference List:


**Provider Training**

The PFR provider was trained to a gold standard fidelity level by the developers of the PFR program over a six-month period, and was an experienced PFR
provider. The provider met with the PFR primary developer weekly during the study for reflective practice to maintain fidelity to the program. Any issues or questions about program delivery were resolved during the weekly reflective meetings.

The Promoting First Relationships training program integrates theory, practice and intervention. Trainers use various learning approaches including video case studies, role-playing, and reflective dialogue. These approaches allow participants to apply the framework directly to issues faced in their work environment. The PFR training program focuses on the following:

- understanding factors that promote a positive parent-infant relationship; using reflective observation, verbal feedback, and reflective questioning;
- discussing social and emotional needs specific to the infant-toddler period;
- discussing care giving qualities and activities that promote a contingent, reciprocal interaction and enhanced behavioral and emotional regulation in the infant (e.g., individualized attention, increased routine and predictability);
- practicing care giving qualities and activities that promote social behavior (e.g., managing distress, offering rituals and routines);
- handling challenging behaviors (e.g., identifying possible causes for challenging behaviors, reframing the behaviors for parents, developing individualized intervention plans); and
- exploring the parent’s own sense of self, emotion regulation and support that influence the care giving environment.

Further details can be found at http://pfrprogram.org/training/.
**Fidelity Coding.**

The PFR provider videotaped 20% of the parent video feedback sessions. The recordings were coded for fidelity. The fidelity coding scheme was developed by the PFR program developer who was also the fidelity coder for the present study. There were three aspects to the fidelity coding. First, the provider’s comments were coded as being either one of the four PFR consultation strategies (positive feedback; positive and instructive feedback; reflective comment or question; or validating, responsive statement) or other. The ratio of PFR reflective consultation strategies to other comments was recorded. Second, the coder categorized the predominant focus of each minute of the segment as being on the dyadic relationship, child, or caregiver. Third, the rater assigned the segment a global rating from "1" - provider used no PFR strategies, and was not relationship-focused, to "5" - provider used all PFR strategies, was solidly focused on the dyadic relationship, and made the reflective dialogue rich and meaningful for the caregiver. To pass fidelity, the provider needed a four to one ratio of PFR consultation strategies to other comments, be predominantly focused on the relationship in 90% of the minutes coded, and receive a 4 or 5 on the global rating. The PFR provider in the present study passed 100% of the fidelity checks.

**Weeks 2-10**

Weeks 2-10 began with reflecting on the prior week's content and focusing on the following areas of child development and parenting: (1) social and emotional needs specific to the infant-toddler period; (2) caregiving qualities and activities that promote a contingent, reciprocal interaction and enhanced behavioral and emotional regulation in the child (e.g., individualized attention, increased routine and predictability); (3) caregiving qualities and activities that promote social behavior
(e.g. strategies for encouraging cooperation and managing distress); (4) challenging behaviors (e.g., identifying possible causes for challenging behaviors, reframing the behaviors for parents, developing individualized intervention plans); (5) the parent’s own sense of self, emotion regulation and support that influence the caregiving environment; and (6) Noticing and responding sensitively to child’s communicative cues. During the course of the 10 sessions, the provider and parent also reviewed up to 15 handouts on topics such as "staying connected during difficult moments; and "qualities that encourage trust and security."

S1.3. Baseline variables

At 6 months there were no group differences prior to intervention (PFR vs. A+M) in the Autism Observation Scale for Infants total score or number of markers (Fs < .54, ps > .42) nor on the Vineland Communication, Daily Living Skills, Social, Motor Skills, or ABC Standard Scores on the Vineland at 6 months (Fs < 2.2, ps > .15). There were no differences between groups in Mullen Adaptive Behavior Scales standard scores on the Gross Motor, Visual Receptive, Fine Motor, Receptive or Expressive Language (Fs < 2.6, ps > .112). This confirms that the groups did not differ in baseline behavioral variables, and thus none of these metrics are used as covariates.

S1.4. Other interventions received

Table S1 gives details of any interventions received by infants in the two high-risk groups during the course of the trial. Some infants received more than one type of intervention; in total 5/14 infants from the A+M group and 6/19 infants from the PFR group received some kind of intervention between 6 and 12 months, and 5/14 in the
A+M group and 3/19 infants from the PFR group received some kind of intervention between 12 and 18 months.

<table>
<thead>
<tr>
<th>Intervention</th>
<th>PFR (total n=19)</th>
<th>A+M (total n=14)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6 to 12 months</td>
<td>12 to 18 months</td>
</tr>
<tr>
<td>Parent-child socialization/support</td>
<td>N=1 (19.8hrs)</td>
<td>N=1 (17.4hrs)</td>
</tr>
<tr>
<td>Parent-child recreation</td>
<td>N=1 (59.4hrs)</td>
<td>N=2 (16.2hrs, 4hrs)</td>
</tr>
<tr>
<td>Occupational Therapy</td>
<td>N=0</td>
<td>N=2 (5.8hrs, 55.6hrs)</td>
</tr>
<tr>
<td>Speech Therapy</td>
<td>N=1 (59.4hrs)</td>
<td>N=1 (23.3hrs)</td>
</tr>
<tr>
<td>Applied Behavioral Analysis</td>
<td>N=2 (23.5hrs, 7.2hrs)</td>
<td>N=1 (70.7hrs)</td>
</tr>
<tr>
<td>Music therapy</td>
<td>N=1 (19.4hrs)</td>
<td>N=1 (23.3hrs)</td>
</tr>
<tr>
<td>Special Education</td>
<td>N=2 (59.4hrs, 1.46hrs)</td>
<td>N=1 (70.7hrs)</td>
</tr>
<tr>
<td>Parenting class</td>
<td>N=1 (19.4hrs)</td>
<td></td>
</tr>
<tr>
<td>Physical Therapy</td>
<td>N=2 (59.4hrs, 1.46hrs)</td>
<td></td>
</tr>
</tbody>
</table>

**Table S1:** Other interventions received by infants in the RCT. Numbers – number of infant in each group who received that type of intervention. In parentheses are total times in hours over each 6-month period received by those infants.
S1.5. Habituation Task

Habituation Stimuli

Stimuli were colored photographs of female faces and toys, measuring 25 cm by 25 cm. Faces were neutral forward facing headshots of Caucasian women. Toys were chosen to be age appropriate (e.g., rattles), horizontal in orientation, symmetrical, forward facing, and free from any social features (e.g., eyes). Four pairs of stimuli were used in each stimulus condition (faces or toys), and were matched on features (toys: size, shape, category, color profile; faces: hair style, skin and hair color, facial size) to ensure findings were not item specific. Stimuli were counterbalanced across participants. Preliminary analysis confirmed group effects did not differ as a function of stimulus set, and so analyses were collapsed across this variable.

Habituation Procedure

At 6, 12 and 18 months, children participated in four habituation experiments, in a two stimulus (faces or toys) by two delay (1 second vs. 1 minute) repeated measures design. Two delays were used to assess whether infants might show difficulties with immediate versus longer-term face or object recognition. Testing was conducted across two different days to reduce possible transfer effects. At each visit, one test involving faces and one test involving toys was presented. At 6 and 18 months, one test at each visit was conducted with the short delay (1 second), and one was conducted with the long delay (1 minute). (For example: Day 1 Face Long delay and Toy Short delay; Day 2 Toy Long delay and Face Short delay). At 12 months both tasks at each visit were conducted with one delay (either long or short). Order of testing for both stimulus and delay was counterbalanced within these restrictions. Parents were asked to refrain from providing verbal or nonverbal cues during the
testing procedure. Preliminary analysis confirmed that group effects did not significantly differ as a function of testing day or order, and so analyses presented were collapsed across these variables.

Infants were seated on their parent’s lap approximately 100 cm from the display; stimuli subtended 14 by 14 degrees of visual angle and were presented on a 46-inch liquid crystal display monitor. A closed-circuit camera was placed underneath the monitor. Two experimenters stood behind a barrier and monitored the infant’s behavior via live feed from the camera. Stimulus presentation was controlled using a custom-built software package (“Looktime”).

During the habituation phase, two experimenters independently measured looking time by pressing a button while the child visually fixated the stimulus. The stimulus was removed from the screen if the child looked away for more than one second (based on online computation of data from the first experimenter). When this occurred, an “attention getter” (a flashing colored square accompanied by a chirping noise) was used to regain the child’s attention to the screen. When the child attended to the screen for longer than one second, the stimulus was re-presented. We used a reorient cue to maximize participant retention and to reduce the potential effect of differences in endogenous orienting on look spacing and habituation times\(^1\); for examples see \(^2\)–\(^5\) and is included in a leading program used to implement habituation protocols\(^6\).

Habituation was defined as having been met when each of two consecutive looks fell below fifty percent of the average of the child’s longest two looks, requiring a minimum of 4 looks (illustrated in Figure 1A). A ‘look’ was defined as visual fixation for greater than one second. These calculations were implemented by the LookTime software. The two longest looks (rather than first look duration) were
chosen as the criterion because around 40% of individual infants produce their peak looks later in the habituation function\textsuperscript{7-9}.

The habituation phase was followed by a delay phase of either one second or one minute. During the delay phase, the child was not shown any stimulus. In the testing phase, the familiar stimulus and a previously unseen stimulus of the same stimulus category were presented in random order. Each stimulus was presented for the duration of one look.

Habituation data was considered valid if: (1) Infants met the habituation criterion; (2) habituation was not judged to be invalid during testing (e.g. child was crying, or eyes couldn’t be seen); and (3) look coding was considered reliable as assessed by calculating the intra-class correlation coefficient between the first and second experimenter coding for all infants. If the correlation was less than $r=0.8$, the video recording was re-coded off-line by trained coders.

**Habituation data processing**

As children participated in two face habituation experiments and two toy habituation experiments at each time-point, summary values (e.g. peak look duration, total time to habituate) were averaged across the two experiments in each condition to provide a more stable characterization of individual differences\textsuperscript{10,11,12}. If a child had only one valid data point for either the toy or the face condition, this data point alone was included in the analysis. This enabled us to maximize the number of children included in the final analysis. Preliminary tests revealed no significant effects of the number of data points included in the analysis for each child (all $ps > 0.1$); this variable was excluded from further analyses.

**Dishabituation**
In order to establish evidence of recognition memory for the familiar stimulus at each delay, the duration of looking to the novel stimulus was compared to the duration of the last look during habituation with a repeated measures analysis. If the last look was significantly shorter than the look to the novel stimulus, dishabituation was inferred. This would indicate detection of the difference between the familiar stimulus and the novel stimulus.

**Final sample (low risk control groups)**

At 6 months the final sample of infants who provided data for both the face and toy habituation included 36 of 36 infants in the Low-Risk\textsubscript{Long} group and 51 of 51 infants in the Low-Risk\textsubscript{Cross} group. At 12 months, 35 of 36 infants in the Low-Risk\textsubscript{Long} group and 53 of 54 infants in the Low-Risk\textsubscript{Cross} group provided valid data for both toy and face habituation. At 18 months, 32 of 36 infants in the Low-Risk\textsubscript{Long} group provided valid data for both toy and face habituation conditions at 18 months.

**Final sample (intervention analysis)**

To investigate change from 6 to 12 months, 10 of 14 infants in the A+M group and 16 of 19 infants in the PFR group provided valid data for both toy and face habituation conditions at both 6 and 12 months. The data from the 7 children that did not contribute data was due to: 2 children (1 A+M and 1 PFR) provided data for both conditions at 6 months only; 2 children provided data for both conditions at 12 months only (1 A+M and 1 PFR); 1 child (PFR) had data on the face condition at 12 months only; 2 children (both A+M) had complete data at 12 months, but data from one condition at 6 months (one face, one toy condition). Children with some cross-sectional data were included in condition-specific cross-sectional analyses at the appropriate time-point to maximize retention.
To investigate change from 6 to 18 months, 8 of 14 infants in the A+M group and 13 of 19 infants in the PFR group provided valid data for both toy and face habituation conditions at both 6 and 18 months and were included in analyses of longitudinal change. Twelve children did not provide data due to: 7 children (3 A+M and 4 PFR) provided data for both conditions at 6 months only; 1 child (A+M) had complete data at 18 months, but data from only the face condition at 6 months; 1 child (A+M) had data from only the toy condition at both 6 and 12 months; 3 children (n=1 A+M, n=2 PFR) had no valid data at either 6 or 18 months. Children with some cross-sectional data were included in condition-specific cross-sectional analyses at the appropriate time-point to maximize retention.

S1.6. EEG theta power to social and non-social videos

Stimuli

Stimuli were a set of videos of social stimuli (4 vignettes of women telling nursery rhymes; each 15 seconds) and non-social stimuli (4 dynamic toys such as a ball dropping down a chute; each 15 seconds). Video sets were presented in random order and repeated twice (e.g., SNSN). Stimuli were presented on a 23-inch monitor using Matlab with the infant seated 65 cm from the monitor.

Procedure

EEG was recorded continuously throughout the session using NetStation 4.3, with a concurrent video record of the infant’s behavior time-locked to the EEG record. The social set and the non-social set order was randomized. The video sets were repeated twice. EEG was recorded from 128 electrodes using the Geodesic Sensor Net (Electrical Geodesics, OR), on Net Station 4.3 data acquisition software.
A 128 lead Geodesic sensor net was dipped into KCl electrolyte solution and placed on the infant’s head and fitted. Electrodes were arranged to symmetrically cover the scalp from nasion to inion and left to right ears. Impedances were typically < 50 kOhm. EEG was recorded with reference to the vertex (Cz) electrode, amplified, and analog filtered (elliptical) between .1 and 100 Hz. Signals were digitized at 500 samples/second. During testing, the infants were monitored for eye and head movements.

Using Net Station 4.3, EEG data were filtered (.1 to 100Hz) to remove high frequency noise. EEG was segmented into 1-second segments. EEG data were manually edited to remove segments with artifact due to eye movements or motion, and to reject electrodes with a preponderance of noise resulting from poor electrode contact with the scalp. An amplitude threshold criterion was applied to reject electrodes that exceeded +/- 200 µV during a trial, and if any given electrode was rejected in more than 25% of trials, it was eliminated entirely for that participant. EEG data were analyzed for spectral power using Matlab v2010a. Averaged power spectra for each subject were visually inspected at all electrodes for artifact contamination. The remaining data was de-trended and processed in a Fast Fourier transform, and the data was again excluded for movement artifact. Power values were then averaged across artifact free segments and electrodes within topographical groups and natural logs were calculated to reduce skew.

In order to calculate the frontal power indices, natural log (ln) 4- to 6-Hz theta and 6 to 9Hz alpha power data from the midfrontal (approximate electrode locations F3/F4; see Figure S1) were used (Marshall, Bar-Haim, & Fox, 2002). Infants were only included in the analyses if they provided at least 5 artifact-free segments per condition.
**Final sample (low risk control groups).**

At 6 months, 34 of 36 infants in the Low Risk\textsubscript{Long} group and 102 of 114 infants in the Low Risk\textsubscript{Cross} group provided sufficient data in the social and nonsocial conditions at 6 months. At 12 months, 30 of 36 infants in the Low Risk\textsubscript{Long} group and 97 of 104 infants in the Low Risk\textsubscript{Cross} group provided sufficient data in the social condition and nonsocial conditions at 12 months. At 18 months, 31 of 36 infants in the Low Risk\textsubscript{Long} group provided sufficient data for both social and nonsocial conditions at 18 months.

**Final Sample (intervention analysis)**

Analysis of change from 6 to 12 months included 15 of 19 infants in the PFR group and 9 of 14 infants in the A+M group who had sufficient data for both social and nonsocial conditions at both ages. A further three infants (n=2 PFR, n=1 A+M) had sufficient data at 12 but not 6 months, and five infants (n=3 A+M and n=2 PFR) had sufficient data at 6 but not 12 months. Children with some cross-sectional data were included in condition-specific cross-sectional analyses at the appropriate time-point to maximize retention. One infant had no data at either time-point (A+M).

Analysis of change from 6 to 18 months included 15 of 19 infants in the PFR group and 12 of 14 infants in the A+M group who had sufficient data for both social and nonsocial conditions at both ages. A further two infants (n=1 PFR, n=1 A+M) had data at 18 but not 6 months, and three infants (n=3 PFR) had data at 6 but not 18 months. Children with some cross-sectional data were included in condition-specific cross-sectional analyses at the appropriate time-point to maximize retention. One infant had no data at either time-point (A+M).

S1.7. Event-Related Potential (ERP) Task
Event-related potential data was collected at 6, 12 and 18 months, but the number of children who provided data at each age is low. Thus analyses of ERP data are restricted to cross-sectional group comparisons.

**Stimuli**

Stimuli included 100 digital photographs of faces (including both internal and external features) and objects (toys) were presented. Face stimuli were chosen to reflect the ethnicity of the local community (86% Caucasian, 8% Asian and 6% African-American); gender was balanced. Toys were photographs of age-appropriate toddlers’ “favorite” toys that did not have a face and were oriented vertically to match faces in size and width, as depicted in Figure 2A. Stimulus frames were 336 pixels wide by 420 pixels high and were presented for 500 msec on an LCD monitor 65 cm from the child at a size of 18 cm by 11 cm, subtending a visual angle of 16 by 10 degrees.

**ERP procedure**

ERPs were recorded from 128-channel Geodesic sensor nets recorded with reference to the vertex. Data was recorded at 500 Hz, with amplification set at 1000x, and band-pass filtering at 0.1 and 100 Hz. Children were presented with a series of 2300-2800 msec trials consisting of: 100 msec baseline, 500 msec stimulus presentation, 1200 msec post-stimulus recording period; 500 - 1000 msec randomly jittered ITI. Testing was terminated when the child had attended to 100 of each of the stimulus types, or when the child was no longer attending. Offline, data were low pass filtered at 20 Hz and segmented into 1800 msec epochs. Artifact detection was accomplished with both automatic artifact-detection software (NetStation 4.3) and through hand-editing. During hand-editing, files were labeled by subject number with no accompanying information about risk status. Trials were rejected if the child did
not attend to the picture (recorded online by a trained observer), if the signal amplitude exceeded 250 µV, if electro-ocular or muscular artifact occurred, or if there was significant drift. Data was re-referenced offline to the average reference and trials were corrected with respect to the 100 msec pre-stimulus baseline period.

Posterior Temporal left and right regions (Figure S2) and components of interest were defined with respect to the previous literature, and inspection of the grand average waveform. These regions substantially overlap those used in previous work with children with ASD\textsuperscript{13–17}. We analyzed P400 peak amplitude and latency because these measures have been sensitive to atypicalities in infants with later ASD\textsuperscript{16,18} and children with ASD\textsuperscript{17} in previous work. Peaks were identified for each electrode using automatic peak detection software, and verified by visual inspection. Peaks were defined as the most positive (P400) point of a deflection between 300 and 900 msec (P400), and the peak had to be present in at least 2/6 electrodes in a group. Peak amplitude and latency values were averaged across regions.

**ERP analysis strategy**

We used repeated measures ANOVAs on P400 latency and amplitude by condition (face, object) and group (High-Risk\textsubscript{A&M}, High-Risk\textsubscript{PFR}). Where interactions were significant, follow-up ANOVAs were used to clarify effects.

**Final sample (low risk control groups)**

At 6 months 19 of 36 infants in the Low Risk\textsubscript{Long} group and 50 of 114 infants in the Low Risk\textsubscript{Cross} group provided sufficient data for analysis. Data loss was due to the infant failing to attend to sufficient trials (n=6 Low Risk\textsubscript{Long}; n= 17 Low Risk\textsubscript{Cross}), poor data quality (n=10 Low Risk\textsubscript{Long}; n= 46 Low Risk\textsubscript{Cross}), and equipment difficulties (n=1 Low Risk\textsubscript{Long}; n= 1 Low Risk\textsubscript{Cross}). Mean numbers of visually attended trials were: Low Risk\textsubscript{Long} Face M=42.3 SE 2.7, Toy M=40.5 SE 3.0; Low Risk\textsubscript{Cross} Face
M=42.3 SE = 1.7, Toy M=42.6 SE = 1.7. Mean number of trials included in the ERP analysis were: Low Risk\textsubscript{Long} Face M=19.4 SE 1.4, Toy M=19.6 SE 1.7; Low Risk\textsubscript{Cross} Face M=19.2 SE = .9, Toy M=19.3 SE = .9.

For 12 months 22 of 36 infants in the Low Risk\textsubscript{Long} group and 59 of 104 infants in the Low Risk\textsubscript{Cross} group provided sufficient data for analysis at 12 months. Data loss was due to the infant failing to attend to sufficient trials (n=6 Low Risk\textsubscript{Long}; n= 19 Low Risk\textsubscript{Cross}), poor data quality (n=5 Low Risk\textsubscript{Long}; n= 23 Low Risk\textsubscript{Cross}), and refusal to wear the EEG net (n=3 Low Risk\textsubscript{Long}; n= 3 Low Risk\textsubscript{Cross}). Mean numbers of visually attended trials were: Low Risk\textsubscript{Long} Face M=43.9 SE 2.3, Toy M=43.1 SE 2.4; Low Risk\textsubscript{Cross} Face M=42.6 SE = 1.4, Toy M=43.3 SE = 1.4. Mean number of trials included in the ERP analysis were: (Low Risk\textsubscript{Long} Face M=20.9 SE 2.1, Toy M=20.3 SE 2.2; Low Risk\textsubscript{Cross} Face M=21.5 SE = 1.0, Toy M=21.1 SE = .9.

For 18 months, 26 of 36 infants in the Low Risk\textsubscript{Long} group provided sufficient data for analysis at 18 months. Data loss was due to the infant failing to attend to sufficient trials (n=1 Low Risk\textsubscript{Long}), poor data quality (n=5 Low Risk\textsubscript{Long}), refusal to wear the EEG net (n=2 Low Risk\textsubscript{Long}), and not attending the EEG visit (n=2 Low Risk\textsubscript{Long}). Mean numbers of visually attended trials were: Low Risk\textsubscript{Long} Face M=48.9 SE 2.1, Toy M=47.9 SE 2.1). Mean number of trials included in the ERP analysis were: (Low Risk\textsubscript{Long} Face M=29.5 SE 2.3, Toy M=26.9 SE 2.2.

**Final sample (intervention analysis)**

For 6 months, 8 of 19 infants in the PFR group and 9 of 14 infants in the A+M group provided sufficient data for analysis at 6 months. Data loss was due to the infant failing to attend to sufficient trials (n=2 PFR; n= 1 A+M), poor data quality (n=7 PFR, n= 4 A+M), and equipment difficulties (n=2 PFR). This attrition rate is comparable to other work in this area\textsuperscript{19}. There were no significant differences in
attended trials ($PFR$ Face $M=36.6$ SE $5.8$, Toy $M=39.3$ SE $5.1$; $A+M$ Face $M=38.7$ SE $6.1$, Toy $M=39.0$ SE $5.5$; $F (1,16) = 0.02$, $p = 0.90$); or trials included in the ERP analysis ($PFR$ Face $M=19.6$ SE $2.2$, Toy $M=19.9$ SE $2.0$; $A+M$ Face $M=15.4$ SE $1.9$, Toy $M=15.2$ SE $1.7$; $F (1,16) = 3.1$, $p = 0.1$) between the groups.

For 12 months, 7 of 19 infants in the $PFR$ group and 6 of 14 infants in the $A+M$ group provided sufficient data for analysis at 12 months. Data loss was due to the infant failing to attend to sufficient trials ($n=6$ PFR, $n=3$ A+M), poor data quality ($n=5$ PFR, $n=3$ A+M), refusal to wear the EEG net ($n=1$ PFR and $n=1$ A+M), and one family (A+M) who declined to attend the EEG visit. This attrition rate is comparable to other work in this area\textsuperscript{16,18}. There were no significant differences in attended trials ($PFR$ Face $M=33.4$ SE $5.3$, Toy $M=32.3$ SE $5.9$; $A+M$ Face $M=47.1$ SE $5.7$, Toy $M=41.7$ SE $5.5$; $F (1,12) = 2.2$, $p = 0.17$); or trials included in the ERP analysis ($PFR$ Face $M=16.7$ SE $3.3$, Toy $M=18.6$ SE $3.5$; $A+M$ Face $M=26.1$ SE $3.4$, Toy $M=22.0$ SE $3.8$; $F (1,12) = 1.7$, $p = 0.22$) between the groups.

For 18 months, 12 of 19 infants in the $PFR$ group and 9 of 14 infants in the $A+M$ group provided sufficient data for analysis at 18 months. Data loss was due to the infant failing to attend the ERP visit ($n=2$ PFR), failure to attend to sufficient trials ($n=1$ A+M), poor data quality ($n=4$ PFR, $n=3$ A+M) and refusal to wear the EEG net ($n=1$ PFR and $n=1$ A+M). This attrition rate is comparable to other work in this area\textsuperscript{20}. At 18 months there were no significant differences between groups in attended trials ($PFR$ Face $M=39.2$ SE $3.6$, Toy $M=39.5$ SE $3.2$; $A+M$ Face $M=49.9$ SE $3.6$, Toy $M=49.9$ SE $3.8$; $F (1,20) = 2.9$, $p = 0.1$); or in trials included in the ERP analysis ($PFR$ Face $M=20.5$ SE $2.3$, Toy $M=20.0$ SE $2.3$; $A+M$ Face $M=24.4$ SE $2.6$, Toy $M=23.1$ SE $2.7$; $F (1,20) = 1.2$, $p = 0.29$) between the groups.
S2. Supplementary Results

<table>
<thead>
<tr>
<th>Low Risk&lt;sub&gt;Cross&lt;/sub&gt;</th>
<th>12 months</th>
<th>18 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>378.2 (1.2)</td>
<td>N/A</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Low Risk&lt;sub&gt;Long&lt;/sub&gt;</th>
<th>12 months</th>
<th>18 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>377.8 (2.5)</td>
<td>570.2 (3.5)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PFR</th>
<th>12 months</th>
<th>18 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>386.9 (6.8)</td>
<td>565.3 (4.6)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>A+M</th>
<th>12 months</th>
<th>18 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>374.5 (2.1)</td>
<td>557.6 (3.4)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANOVA by treatment group</th>
<th>12 months</th>
<th>18 months</th>
</tr>
</thead>
<tbody>
<tr>
<td>F(1,26) = 1.60, p = 0.22</td>
<td>F(1,19) = 0.34, p = 0.6</td>
<td></td>
</tr>
</tbody>
</table>

Table S2: Chronological age in days at the 12 and 18-month visits.

S2.1. Comparisons between high risk infants and longitudinally assessed low-risk controls.

We repeated cross-sectional analyses reported in the main document with the inclusion of the longitudinally-assessed low risk control group, to identify whether there were any differences between treated and untreated high-risk infants and the normative controls at any time-point.

**Habituation**

Cross sectional comparison using ANOVA on face and object total habituation times by Group (PFR, A+M, Controls<sub>long</sub>) at 6, 12 and 18 months separately showed a significant group difference in face habituation times at 18 months (Face: $F(2,54) = 5.36$, $p = 0.008$, $\eta^2 = 0.17$); no other comparisons reached significance ($ps > 0.09$). Post-hoc Bonferroni-corrected t-tests indicated that this was due to shorter habituation times in PFR vs A+M groups ($p=0.006$); there were differences between the A+M group and the low-risk controls at the trend level such
that the A+M group had prolonged habituation times to faces (p=0.077); there were no significant difference between the PFR group and low-risk controls (p = 0.29).

**EEG**

Cross-sectional analyses indicated no significant effects of group at any time point (ps > 0.05).

**ERP**

*P400 amplitude:* Cross-sectional analyses indicated a significant interaction between Group (A+M, PFR, Control<sub>Long</sub>) and Stimulus (face, object) on P400 amplitude at 12 months ($F(2,32) = 4.06, p = 0.027$). Post-hoc analyses split by Stimulus did not reveal significant effects. Post-hoc analyses split by Group revealed that in addition to results reported in the main text, the Control<sub>Long</sub> group showed a significant condition effect such that responses to objects were larger than responses to faces ($F(1,21) = 7.94, p = 0.01$).

There were no other significant effects of group at any timepoint (ps > 0.05).

*P400 latency:* Cross-sectional analyses indicated a significant interaction between Group (A+M, PFR, Control<sub>Long</sub>) and Stimulus (face, object) on P400 latency at 18 months ($F(2,44) = 5.78, p = 0.006$). Post-hoc analyses split by Stimulus did not reveal significant effects. Post-hoc analyses split by Group revealed that in addition to results reported in the main text, the Control<sub>Long</sub> group showed a significant condition effect such that responses to objects were faster than responses to faces ($F(1,23) = 4.79, p = 0.039$).

There were no other significant effects of group at any time-point (ps > 0.05).
Figure S1: Location of electrodes used in the frontal theta and P400 analyses.
Figure S2: Change in habituation time from 6 to 12 and 6 to 18 months for faces and objects.

Scores on the y axis represent difference scores; negative scores indicate that habituation times dropped between 6 and 12/18 months. PFR are infants with an older sibling with ASD who received the Promoting First Relationships intervention; A+M are infants with an older sibling with ASD who received Assessment and Monitoring; Controls - Long are longitudinally assessed infants with no family history of ASD.
Figure S3: Change in EEG power from 6 to 12 and 6 to 18 months for social and nonsocial videos.

Scores on the y axis represent difference scores; positive scores indicate that EEG power increased between 6 and 12/18 months. PFR are infants with an older sibling with ASD who received the Promoting First Relationships intervention; A+M are infants with an older sibling with ASD who received Assessment and Monitoring; Controls - Long are longitudinally assessed infants with no family history of ASD.
S.4. Supplementary References


