Developmental Science

Developmental Science (2013), pp 1–13 DOI: 10.1111/desc.12112

PAPER

The bonnie baby: experimentally manipulated temperament affects perceived cuteness and motivation to view infant faces

Christine E. Parsons,[1,2] Katherine S. Young,[1,2] Ritu Bhandari,[3] Marinus H. van Ijzendoorn,[3] Marian J. Bakermans-Kranenburg,[3] Alan Stein[1] and Morten L. Kringelbach[1,2]

1. *Department of Psychiatry, University of Oxford, UK*
2. *Centre of Functionally Integrative Neuroscience, Aarhus University, DK*
3. *Centre for Child and Family Studies Leiden University, Leiden, The Netherlands*

Abstract

Attractive individuals are perceived as having various positive personality qualities. Positive personality qualities can in turn increase perceived attractiveness. However, the developmental origins of the link between attractiveness and personality are not understood. This is important because infant attractiveness ('cuteness') elicits caregiving from adults, and infant personality ('temperament') shapes caregiving behaviour. While research suggests that adults have more positive attitudes towards cuter infants, it is not known whether positive infant temperament can increase the perception of infant cuteness. We investigated the impact of experimentally established infant temperament on adults' perception of cuteness and desire to view individual faces. At baseline, adults rated the cuteness of, and keypressed to view, images of unfamiliar infants with neutral facial expressions. Training required adults to learn about an infant's 'temperament', through repeated pairing of the neutral infant face with positive or negative facial expressions and vocalizations. Adults then re-rated the original neutral infant faces. Post-training, there were significant changes from baseline: infants who were mostly happy were perceived as cuter and adults expended greater effort to view them. Infants who were mostly sad were not perceived as cuter and adults expended less effort to view them. Our results suggest that temperament has clear consequences for how adults perceive 'bonnie' babies. Perception of infant cuteness is not based on physical facial features alone, and is modifiable through experience.

Introduction

We instinctively make judgements about others based on physical appearance. An attractive face can lead viewers to make positive inferences about personality; a bias termed the 'what is beautiful is good' stereotype. In almost every context, attractive people fare better than unattractive people (Eagly, Makhijani, Ashmore & Longo, 1991; Langlois, Kalakanis, Rubenstein, Larson, Hallam & Smoot, 2000), be it in mating success, job applications or election results. Although studied to a far lesser extent, attractiveness in children also appears to be a determinant of treatment by unfamiliar adults. Adults tend to rate the misbehaviour of attractive children less harshly than that of less attractive children (Berkowitz & Frodi, 1979; Dion, 1972) and provide more positive academic evaluations (Parks & Kennedy, 2007; Ritts, Patterson & Tubbs, 1992). For infants, research is more limited, but one correlational study has demonstrated that mothers of more attractive offspring have more positive attitudes and affectionate behaviours towards their infants compared with mothers of less attractive offspring (Langlois, Ritter, Casey & Sawin, 1995).

Infant attractiveness also correlates with judgements of infant health, pleasantness, and adoption preferences (Karraker & Stern, 1990; Maier, Holmes, Slaymaker & Reich, 1984; Ritter, Casey & Langlois, 1991; Stephan & Langlois, 1984; Volk & Quinsey, 2002). Perhaps the most convincing evidence for the importance of infant facial features in eliciting care comes from instances of facial

Address for correspondence: Morten L. Kringelbach, University Department of Psychiatry, Warneford Hospital, Oxford OX3 7JX, UK; e-mail: Morten.Kringelbach@psych.ox.ac.uk

anomaly, signalling compromised health. Cleft lip in infants has been associated with disruption of parental care (Field & Vega-Lahr, 1984; Murray, Hentges, Hill, Karpf, Mistry, Kreutz, Woodall, Moss & Goodacre, 2008), particularly in the case of severe disfigurement (Murray *et al.*, 2008). Adults who are not parents have also been shown to react negatively to both specific and global changes to the infant face, such as in cleft lip, fetal alcohol syndrome and prematurity (Frodi, Lamb, Leavitt & Donovan, 1978; Parsons, Young, Parsons, Dean, Murray, Goodacre, Dalton, Stein & Kringelbach, 2011b; Waller, Volk & Quinsey, 2004).

The strength of preference for attractive over less attractive faces is also exemplified by experimental evidence showing that adults will expend greater effort to view more attractive faces compared to less attractive faces (Aharon, Etcoff, Ariely, Chabris, O'Connor & Breiter, 2001; Hahn, Xiao, Sprengelmeyer & Perrett, 2013; Parsons, Young, Kumari, Stein & Kringelbach, 2011a). Neuroimaging studies have demonstrated a biological basis for this preferential response to attractive faces, be they young or old. Brain regions that typically respond to reward, such as money or pleasant sensory stimuli, also respond more to attractive compared to unattractive adult faces (e.g. Aharon *et al.*, 2001; O'Doherty, Winston, Critchley, Perrett, Burt & Dolan, 2003). One reward region, the orbitofrontal cortex, has been shown to respond rapidly to typical infant faces (Kringelbach, Lehtonen, Squire, Harvey, Craske, Holliday, Green, Aziz, Hansen, Cornelissen & Stein, 2008) but not to infant faces with a structural abnormality (Parsons, Young, Mohseni, Woolrich, Thomsen, Joensson, Murray, Goodacre, Stein & Kringelbach, 2013a). Activity in another key reward region, the nucleus accumbens, can also be modulated by subtle graded manipulation of the cuteness of infant faces (Glocker, Langleben, Ruparel, Loughead, Valdez, Griffin, Sachser & Gur, 2009). The predominant interpretation of these findings is that differential activation of these brain regions occurs because attractive faces are rewarding (e.g. Said, Haxby & Todorov, 2011).

Our bias in favour of attractive individuals is inherently unfair but appears to be deep-rooted, and has an effect even early in life. Therefore, it is important to understand its implications and, given its apparent ubiquity, how to alter perceptions of attractiveness for the better. By tradition, we tend to explain facial attractiveness in terms of physical features. For an infant, these features include big eyes, round cheeks, a large forehead and a small nose (Lorenz, 1943). For adults, these are gender-dependent, but symmetry and how close the face is to the population average are considered important markers of attractiveness. 'Beautiful faces', those that come closest to the prototypical attractive structure, are strongly preferred. For faces of any age, ratings of attractiveness, or the infant-appropriate 'cuteness', are remarkably consistent across viewers (for review, see Langlois *et al.*, 2000) and positively correlate with favourable judgements.

While the link between attractiveness and positive personality attributions appears robust, what about in the other direction? Does personality have an impact on the perception of beauty? There is some evidence that adults' perceptions of the attractiveness of other adults can be altered by personality information and familiarity. For instance, perception of physical attractiveness, based on evaluations of known individuals in high school yearbooks, can be influenced both by familiarity and personality traits (Kniffin & Wilson, 2004). Providing positive personality information can also decrease the threshold for what is considered attractive, while negative personality information can increase the threshold (Swami, Furnham, Chamorro-Premuzic, Akbar, Gordon, Harris, Finch & Tovée, 2010). However, the relationship between personality, or 'temperament', and perceived cuteness has not been investigated in infants. This is of interest because temperament can dictate how infants respond to the environment (for review, see Van IJzendoorn & Bakermans-Kranenburg, 2012) and infant cuteness is hypothesized to elicit nurturing responses in adults (Glocker *et al.*, 2009; Lorenz, 1971; Parsons, Young, Murray, Stein & Kringelbach, 2010; Tinbergen, 1951).

Temperament is thought to render some infants more or less susceptible to environmental influences, be they positive or negative (Belsky, Bakermans-Kranenburg & Van Ijzendoorn, 2007; Ellis, Boyce, Belsky, Bakermans-Kranenburg & Van Ijzendoorn, 2011). While the definitions, interpretations and measures of temperament are manifold, there is a consensus that temperamental characteristics should be present early in life, show moderate stability, and have neurobiological correlates (e.g. Rothbart & Bates, 2006; Zentner & Bates, 2008). It is also generally accepted that the advantages or disadvantages of temperamental characteristics vary depending on context (e.g. deVries, 1984). That is, there is no 'ideal' temperament independent of environmental factors (Van IJzendoorn & Bakermans-Kranenburg, 2012).

One aspect of infant temperament, 'emotionality', has been shown to impact upon the quality of parental care and child outcome, in the presence of other risk factors. Negative emotionality has been associated with less supportive parenting in low socioeconomic status (SES) families in a recent meta-analysis (but not in high SES families; Paulussen-Hoogeboom, Stams, Hermanns & Peetsma, 2007). Persistent infant crying, lasting more

than several months, has also been associated with greater hyperactivity and academic difficulties in childhood (Wolke, Rizzo & Woods, 2002). Findings have been more variable in studies where infant temperament is measured using observation rather than parental report (Paulussen-Hoogeboom *et al.*, 2007). This measurement issue is challenging; parent and observer measures of infant temperament do not always correspond (e.g. Seifer, Sameroff, Barrett & Krafchuk, 1994) and caregiver negative emotion can influence ratings of infant temperament (Youngstrom, Izard & Ackerman, 1999). It has therefore been argued that temperament should be measured in observational settings because the concept of temperament refers to behavioural style and not to parental perceptions (e.g. Kagan, 2007, 2009).

In this study, we designed a novel paradigm to create a sense of the emotional 'temperament' of a number of previously unfamiliar infants. As in real life, temperament was graded, such that some infants were happy most of the time, some were sad most of the time, and others fell somewhere in the middle. We examined the impact of this experimentally manipulated temperament on perception of infant cuteness and also willingness to work to view the infant. These two measures were chosen because adults have been shown to be sensitive to subtle computer-manipulations of infant cuteness, as well as real-world differences in infant facial configuration (Parsons *et al.*, 2011a; Sprengelmeyer, Perrett, Fagan, Cornwell, Lobmaier, Sprengelmeyer, Aasheim, Black, Cameron, Crow, Milne, Rhodes & Young, 2009). In addition, adults will work more to see infants with a 'cuter' face structure compared with less 'cute' faces (Parsons *et al.*, 2011a; Sprengelmeyer, Lewis, Hahn & Perrett, 2013).

We hypothesized that 'positive' infant temperament would be associated with increased ratings of infant cuteness and desire to view, whereas 'negative' infant temperament would be associated with decreased cuteness and desire to view.

Methods

Participants

Participants were 47 healthy women aged between 17 and 27 (M = 19.81 years, SD = 1.59), none of whom had children. All participants were undergraduate students from the Child and Family Studies Department, University of Leiden. Thirty-five of the participants reported having some experience caring for infants (baby-sitting). All participants reported no current or past neurological/psychological disorders, were non-smokers and reported no recreational drug use for at least 6 months. The majority of participants (83%) reported being in the luteal menstrual phase at the time of the study. The study was approved by the ethics committee at the Leiden University Medical Center.

Procedure

The experiment consisted of three stages: baseline measures of participants' responses to the infant faces, the experimental manipulation of infant temperament, and post-manipulation measures of participants' responses to the infant faces (see Figure 1). In the first stage, participants were asked to rate a series of images of infant faces for 'cuteness'. Next, they rated the same

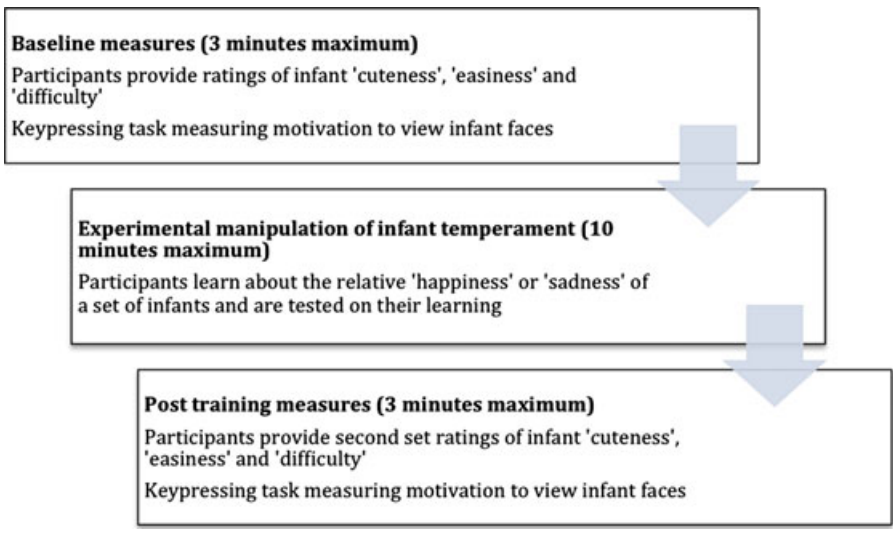

Figure 1 *The three stages of the experimental procedure and their timings.*

images for two characteristics of infant temperament: how 'difficult' and how 'easy' they were perceived to be. They then completed a key press task (referred to elsewhere as a 'wanting' task; e.g. Parsons *et al.*, 2011a). In this task, participants were told to indicate the length of time that they would like to see the images appear on screen by key pressing 'up' (retain) or 'down' (remove; Parsons *et al.*, 2011a). The actual time that the picture remained onscreen did not vary.

In the second stage of the experiment (approximately one hour later), participants performed a novel probabilistic infant social reward task. This task was based on a widely used learning paradigm using three stimulus pairings (Frank, Seeberger & O'Reilly, 2004; Kringelbach & Rolls, 2003) which has also been used with adult faces (Kringelbach & Rolls, 2003) and is described in detail below. Over the course of this task, participants learned about the 'temperament' of different infants. Participants were required to select one infant image from a concurrently presented pair, and were provided with feedback in the form of a change of facial expression and an equivalent vocalization (either 'happy' or 'sad').

Finally, in the third stage, participants were asked to rate the cuteness and temperament of each infant again and complete the 'wanting' task again. Identical instructions were provided for the baseline and post-manipulation measures. All tasks were programmed and performed using Presentation software (Version 14.4 Neurobehavioral Systems, Inc., www.neurobs.com).

Stimuli

All infant face images and vocalization recordings were taken from two standardized databases previously described (for faces, see Kringelbach *et al.*, 2008; Parsons, Young, Parsons, Stein & Kringelbach, 2012; for vocalizations, see Parsons, Young, Joensson, Brattico, Hyam, Stein, Green, Aziz & Kringelbach, 2013b; Young, Parsons, Stein & Kringelbach, 2012). In brief, visual and audio recordings in these databases were collected from infants filmed in their own homes during a play and feeding session with a caregiver. An independent sample of adults rated these stimuli on measures of valence and how they made the rater feel (e.g. 'please rate your mood after listening to this sound'). Categorization of stimuli as 'happy', 'sad' or 'neutral' was based on these ratings.

Visual stimuli were 'happy' (smiling), neutral and 'sad' (crying) faces of six different infants. All of the 'neutral' infant faces had approximately similar valence ratings (scale from −4, very negative affect, to +4, very positive affect; $M = -0.3$, $SD = 0.6$). For each infant, the most positive ($M = 2.7$, $SD = 0.4$) and most negative expressions ($M = -1.8$, $SD = 0.5$) available were used for the 'feedback' during training. Infants were all full-term, healthy, Caucasian and aged between 6 and 8 months at the time of recording. Parental permission was provided for the use of these images for research purposes, which was also approved by the Oxford Research Ethics Committee.

In order to reduce any confounding effects of stimulus gender, we had an independent sample of adult females ($n = 40$) rate faces from a larger set of 13 (Kringelbach *et al.*, 2008) as 'male', 'female', or 'cannot tell'. These ratings were then used to select six faces such that there were two faces clearly perceived as female, two as male, and two with ambiguous ratings. In the 'infant social reward task' (described later), we also used these gender ratings to form same-gender pairings (i.e. both faces in a pair perceived as male). All faces were forward facing, with comparable direction of eye gaze and were matched for size (300 × 300 pixels) and luminosity. Images were presented in greyscale on a 15.3 inch computer monitor.

Six recordings of 'happy' (laughing) and 'sad' (crying) vocalizations were selected. These vocalizations were selected based on ratings from an independent sample of adults to ensure that they unambiguously conveyed these emotions (see Young *et al.*, 2012). Vocal stimuli were free from background noise, matched for peak and average root-mean-square intensity, clipped to 1500 ms and had 150 ms linear rise and fall times (Adobe Audition CS5.5 v4.0). Vocalizations were presented at 50 dB above the hearing threshold for each participant using in-ear earphones (Sennheiser CX300II). For the feedback (visual and auditory), positive and negative vocalizations were paired with positive and negative facial expressions respectively.

Baseline measures: ratings of cuteness and temperament

All participants in the current study rated the neutral facial expressions of each of the six infants on a 'cuteness' scale. These faces were all unfamiliar to participants. Face images were presented in the centre of the screen with a vertical visual analogue scale (VAS) immediately to the right. Participants were asked to 'indicate how cute you find each baby' from 'cute' to 'not cute' on the VAS. The ratings bar started at the midpoint on the scale (halfway between 'cute' and 'not cute') and participants adjusted the height of this bar using the 'up' and 'down' arrows on a standard keyboard. Scores on the VAS ranged from a maximum of 4 (cute) to a minimum of −4 (not cute), with intervals of .0025. This allowed us to obtain sensitive ratings. Each stimulus was presented for 5 seconds and each participant rated each face once. The order of images presented

was randomized between participants. This provided a measure of 'subjective liking' of each image, similar to procedures described in other studies (Parsons *et al*., 2011a; Parsons *et al*., 2011b; Yamamoto, Ariely, Chi, Langleben & Elman, 2009).

Similar to the cuteness ratings, participants also provided ratings of how 'difficult' and how 'easy' each infant was perceived to be. The aim of this was to assess the extent to which participants viewed the infant faces as having different 'temperaments'. That is, these ratings provided a test of the efficacy of the infant social reward task in creating a sense of easy (happy, positive reactivity) and difficult (sad, negative reactivity) infants. This measure of perceived temperament is similar in some respects to reports of infant temperament that might be obtained from a parental questionnaire. Such measures of perceived temperament are important regardless of whether, objectively evaluated, they are correct or not, because they can guide caregiving (e.g. Mäntymaa, Puura, Luoma, Salmelin & Tamminen, 2006).

Face images were presented in the centre of the screen and participants used a VAS to rate how 'difficult' they found each baby (from 'difficult' to 'not difficult') and how 'easy' they found each baby (from 'easy' to 'not easy'). In total, it took participants approximately 2 minutes to complete all of these ratings.

Average cuteness ratings were normally distributed, as confirmed by Kolmogorov-Smirnoff tests ($D(47) = 0.09$, $p = .20$). Temperament ratings were also normally distributed as confirmed by Kolmogorov-Smirnoff tests ('difficulty' rating: $D(47) = 0.12$, $p = .12$, 'easiness' rating: $D(47) = 0.06$, $p = .20$).

Baseline measure: wanting task

In the wanting task, participants were asked to key press to indicate how long they would like to view each image (similar to the procedure described in Parsons *et al*., 2011a). The onscreen appearance was similar to that for the ratings task, with a face image in the centre of the screen and a vertical VAS to the right. The VAS provided a real-time indication of how much key pressing a participant had done (similar to an egg-timer). The actual viewing duration of each image did not vary from trial to trial. However, the amount of key pressing provided a measure of the incentive salience or amount of 'wanting' to view each image, similar to other paradigms (Aharon *et al*., 2001; Parsons *et al*., 2011a). Each stimulus was presented onscreen for 6 seconds and all participants were presented with each face once, lasting a total of 36 seconds. Data from the 'wanting' task were the frequency of 'up' and 'down' key presses, ranging from a minimum of 0 to a maximum of approximately 30.

Due to computer errors, there were two participants with missing data for this task only. After removal of one participant (whose data fell outside the mean plus 3 standard deviations), data were normally distributed, as confirmed by Kolmogorov-Smirnoff tests ($D(44) = 0.13$, $p = .07$). Analyses were conducted including and excluding this outlier and the pattern of results was the same. Results reported here include this outlier.

Infant social reward task

The infant social reward task consisted of a probabilistic learning task, in which participants learned about the 'temperament' of six infants. It was split into two phases: the 'training', where participants learned about each infant, and the 'testing', where the extent of each participant's learning was measured.

Training

The participants' task was to establish which infant from a choice of two was the 'happy' and which was the 'sad' infant through trial and error. Participants selected one of the images and were provided with feedback in the form of a positive or negative facial expression and vocalization. Over repeated trials, participants could then compare how often each infant laughed or cried and work out which infant was the happier of the pair.

Participants were presented with pairs of neutral infant faces and were instructed that, 'in each pair of faces, there is one happy and one sad baby. Like in real life, the happy baby will not always be happy and the sad baby will not always be sad. In each set, your task is to find the happier baby, the one who smiles most often, and continue to always select this baby even if this baby may sometimes appear sad.' For each trial, stimuli were presented in a pair and a total of three pairs were used. The pair combinations were fixed in this phase.

The three pairs varied in the probability of each infant being 'happy' or 'sad'. In the 'easiest' pair, the happy infant laughed on 80% of trials and cried on the remaining 20% of trials. The sad infant in this pair laughed in 20% of trials and cried in the other 80% of trials, if selected. In the second pair, the happy infant laughed 70% of the time and the sad infant laughed 30% of the time. In the final, and hardest to learn pair, the happy infant laughed 60% of the time while the sad infant laughed 40% of the time. Participants only received feedback about the infant that they selected (it was inferred that the infant *not* selected would show the opposite emotion on each trial). In this way, the aim was to establish a sense of graded temperament across the group of infants.

Participants completed two training rounds each consisting of 60 trials. In each training round, every pair of infant was presented 20 times. In one round, participants were asked to select the 'happy' infant in each pair and in the other round to select the 'sad' infant. The order of training rounds (selecting the 'happy' infant or selecting the 'sad' infant) was counterbalanced. By asking participants to deliberately choose the happy or sad infants, we aimed to prevent greater exposure to one individual face in each pair.

At the start of each trial, neutral faces were presented, one on the top half of the centre of a computer screen and the other in the bottom half (see Figure 2a). Participants then selected one image using either the 'up' key (for the top image) or the 'down' key (for the bottom image). Participants were then presented with feedback: the selected image changed to either a happy face accompanied by a laugh, or a sad face accompanied by a cry. The unselected image remained neutral during feedback (see Figure 2b).

In each trial, neutral faces were presented onscreen until participants made a response. After participants made a response, visual feedback was presented immediately for 1.5 sec accompanied by a 1.5 sec vocalization. There was a 500 ms gap between the end of the feedback and the beginning of the next trial during which a red fixation cross was presented in the centre of the screen.

The order of trials during this phase was fully randomized between sessions and participants. The identity of the image from each pair that appeared at the top of the screen was also randomized between trials. Further, the actual identities of which infant face was 80% happy, 70% happy and so on were randomized between participants. This was to exclude the possibility of differences in perceived temperament being related to the appearance of the infants. Depending on the speed of participants' responses, this phase typically lasted between 3 and 5 minutes. In general, participants found learning the contingencies in this task difficult, as evidenced by accuracy rates in the first round of training. For the 80–20% pair, mean accuracy scores were 78% (*SD* = 21%), for the 70–30% pair, 76% (*SD* = 26%) and for the 60–40% pair, 67% (*SD* = 27%). By the second round of training, accuracy rates were higher, but participants were still scoring below 80% on the most difficult pair (60–40% comparison, *M* = 75%, *SD* = 32%). For the other two pairings, mean accuracy rates were at 88% (*SD* = 23%) for the 80–20% pair and at 86% (*SD* = 20%) for the 70–30% pair.

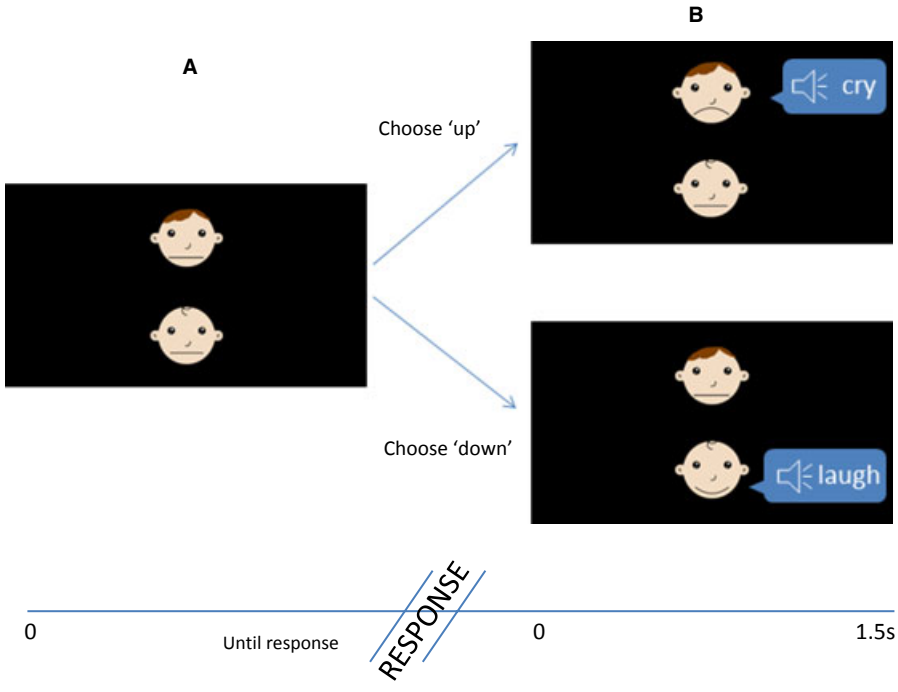

Figure 2 *Example of a single trial in the infant social reward task. (A) Participants are presented with neutral infant faces. (B) After choosing one of the images, participants are presented with audiovisual feedback. Please note that the cartoon face images shown here are for illustration purposes only, the experiment used natural images of infant faces as described in the text.*

Testing

In the 'testing' phase of the task, participants were presented with randomly paired infant images (e.g. 80% with 60% happy infant) and were asked to select the happier of the two infants. This task measured how well each individual learned the previous contingencies and therefore the temperament of the infants.

The set-up of the testing phase was identical to the training phase with a few exceptions. First, no feedback was provided after each trial. Once a participant made a response, there was a 500 ms gap before the next trial was presented, during which a red fixation cross was presented in the centre of the screen. Second, faces were no longer presented just with their specific partner; each face was presented four times with each of the other five faces. This resulted in a total of 60 trials. Finally, participants were instructed to: 'Please choose the face of the baby that "feels" the most happy based on what you have learned during the previous task. If you are not sure which one to pick, just go with your gut feeling.' If a participant failed to make a response within 4.5 sec, the word 'missed' appeared onscreen for 500 ms and the task moved on to the next trial. This phase lasted for a maximum of 5 minutes, but typically lasted around 3 minutes.

Correct responses were counted if participants chose the 'happier' of the two faces (e.g. the 70% happy infant face over the 60% happy infant face). Participants' scores ranged from 40% to 91% correct (M = 71%, SD = 14%).

Post infant social reward task measures: cuteness, temperament, and the wanting task

In the final stage of the experiment, participants completed the cuteness rating, temperament ratings and key press task again. Kolmogorov-Smirnoff tests showed that all rating data were normally distributed (cuteness rating: $D(47) = 0.10, p = .20$; 'difficult' rating: $D(47) = 0.07, p = .20$; 'easiness' rating: $D(47) = 0.11$, $p = .20$). With the outlier (mentioned previously) removed, the wanting task data were also normally distributed ($D(44) = 0.09$, $p = .20$).

Results

Impact of cuteness on temperament

We compared the initial cuteness ratings with temperament ratings to investigate the association between physical face structure and perceived 'personality' characteristics. These ratings were obtained prior to learning about the manipulated temperament of individual infants, so were based solely on facial configuration. A linear regression was used to compare cuteness ratings with perceived temperament. A measure of perceived 'temperament' was obtained by taking an aggregate score of the separate 'easy' and 'difficult' characteristic ratings (which were significantly correlated: $r = -0.76$, $p < .001$). The sign of 'difficult' ratings was reversed and temperament was taken as the mean of the easy rating and the inverted difficult rating. There was a significant relationship between ratings of cuteness and perceived temperament ($R^2 = 0.03$, $p = .003$), indicating a small association between facial features and perceived temperament.

Impact of manipulated temperament on cuteness

To compare the effects of manipulated 'happy' or 'sad' temperament on ratings of infant cuteness, a 2 × 2 repeated measures ANOVA was used. Time (before and after training) and valence (happy and sad) were the factors. There was a significant main effect of time ($F(1, 46) = 16.83$, $p < .001$, $r = 0.52$), with faces rated as cuter after the training task ($M = 1.59$, $SD = 0.81$), compared with before ($M = 1.10$, $SD = 0.72$). There was no significant main effect of infant valence ($F(1, 46) = 1.28$, $p = .26$, $r = 0.16$). There was a significant interaction between time and valence ($F(1, 46) = 10.52$, $p = .002$, $r = 0.43$). Ratings for infants with happy temperament increased more (rated $M = 1.03$, $SD = 1.02$ before training, and $M = 1.90$, $SD = 1.09$ after) than the ratings for infants with sad temperament (rated $M = 1.17$, $SD = 1.19$ before training, and $M = 1.28$, $SD = 1.25$ after). For the 'sad' infants, the change in perceived cuteness was small and the variance in ratings was large. Therefore, the main effect of time appeared to be a result of the change in ratings of the 'happy' infants. Indeed, paired t-tests comparing ratings before and after training (Figure 3) showed that only the ratings for the happy infants increased significantly (for the 80% happy infant, $t(46) = -2.90$, $p = .01$, $r = 0.39$; for the 70% happy infant, $t(46) = -3.03$, $p = .004$, $r = 0.41$, and for the 60% happy infant $t(46) = -3.14$, $p = .003$, $r = 0.42$). All other comparisons were nonsignificant.

Impact of manipulated temperament on desire to view faces

Again, a 2 × 2 repeated measures ANOVA with time and valence as factors was used. There was a significant main effect of valence ($F(1, 44) = 4.95$, $p = .03$, $r = 0.32$) and a significant interaction between time and valence ($F(1, 44) = 21.51$, $p < .001$, $r = 0.57$). There was no significant main effect of time ($F(1, 44) = 0.03$, $p = .87$, $r = 0.03$). On

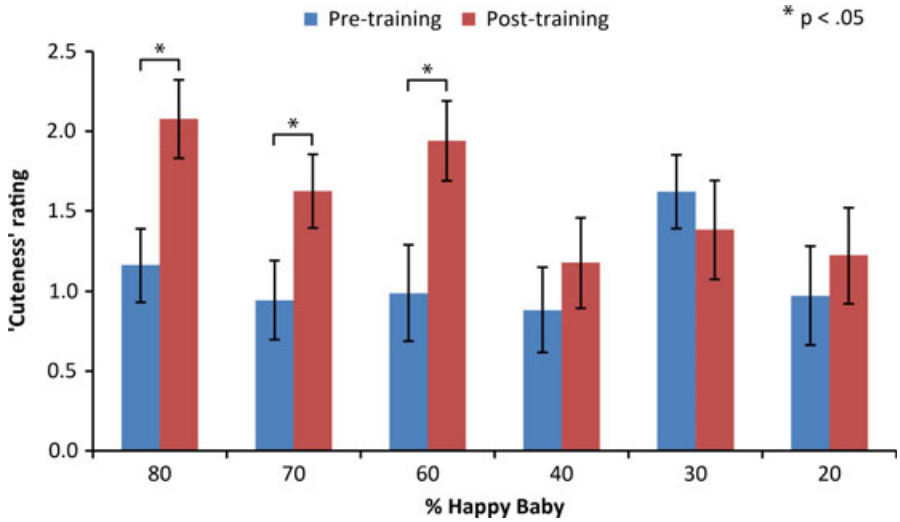

Figure 3 *Cuteness ratings before and after training. 'Happy' infants (80, 70, 60%) were rated as significantly cuter after training. Error bars indicate mean ± standard error.*

average, there were a greater number of key presses for happy infants (M = 5.02, SD = 7.11) compared with sad infants (M = 2.49, SD = 6.41). This effect was largely driven by the difference in key pressing after training. The amount of key pressing for happy infants increased after training (M = 3.29, SD = 6.56 before training, to M = 6.72, SD = 9.48 after training), while the amount of key pressing for sad infants decreased (M = 3.96, SD = 6.97 before training, to M = 0.79, SD = 7.34 after).

Paired t-tests comparing key pressing before and after training for individual faces (Figure 4) showed significant increases for the 80% and 70% happy infants ($t(44) = -2.54$, p = .02, r = 0.36, and $t(44) = -2.15$, p = .04, r = 0.31, respectively). There were also significant decreases in the number of key presses for the 30% and 20% happy infants ($t(44)$ = 2.79, p = .01, r = 0.39, and $t(44)$ = 2.82, p = .01, r = 0.39, respectively).

Effects of manipulated temperament on 'easiness/difficulty' ratings

Again, a 2 × 2 repeated measures ANOVA with time and valence as factors was used. There was a significant main effect of valence ($F(1, 46) = 20.96$, $p < .001$, $r = 0.56$) with the temperament ratings for the happy infants being significantly higher (or 'easier' M = 0.79, SD = 0.83) than for sad infants ($M = -0.05$, SD = 0.66). There was no significant main effect of time ($F(1, 46) = 0.03$, p = .86, r = 0.03), but a significant interaction between time and valence ($F(1, 46)$ = 38.29, p = .001, r = 0.67). The interaction effect showed that this effect of valence was driven by the difference in ratings after training. Temperament ratings for the happy infants increased after training (before training, M = 0.44, SD = 0.87; after training, M = 1.14, SD = 1.10) and ratings for sad infants decreased (more difficult) after training (before training, M = 0.32, SD = 0.72; after training, $M = -0.42$, SD = 0.99; see Figure 5)).

Paired t-tests demonstrated a significant increase in temperament ratings for the 80% happy infant ($t(46) = -2.75$, p = .01, r = 0.38) and the 70% happy infant ($t(46) = -3.28$, p = .002, r = 0.44); and a significant decrease in temperament ratings of the 30% happy infant ($t(46) = 4.97$, $p < .001$, r = 0.59) and the 20% happy infant ($t(46)$ = 3.48, p = .001, r = 0.46).

Linear regression analyses were performed to investigate the impact of temperament on perception of cuteness and desire to view the infant faces. Change scores were calculated for the two measures by subtracting the pre-training ratings from the post-training ratings, and taking absolute values. This allowed us to obtain a measure of change in an individual's response after training, irrespective of whether the change was positive or negative across individual faces.

There was a significant positive relationship between change in temperament ratings and change in cuteness ratings (R^2 = 0.26, $p < .001$). In addition, there was a significant positive relationship between change in temperament ratings and change in number of key presses (R^2 = 0.11, p = .02). There was also a significant positive relationship between change in cuteness and change in number of key presses (R^2 = 0.20, p = .002). As ratings of temperament increased, participants rated faces as cuter and demonstrated increased desire to view them.

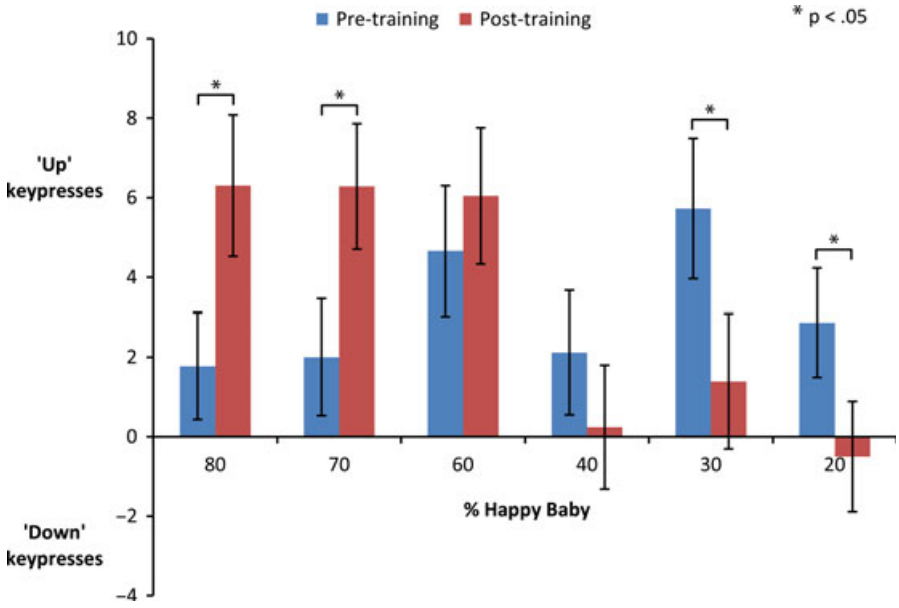

Figure 4 *Participants key pressed significantly more after training compared with before for the two happiest infants. Key pressing decreased significantly after training compared with before for the two least happy infants.*

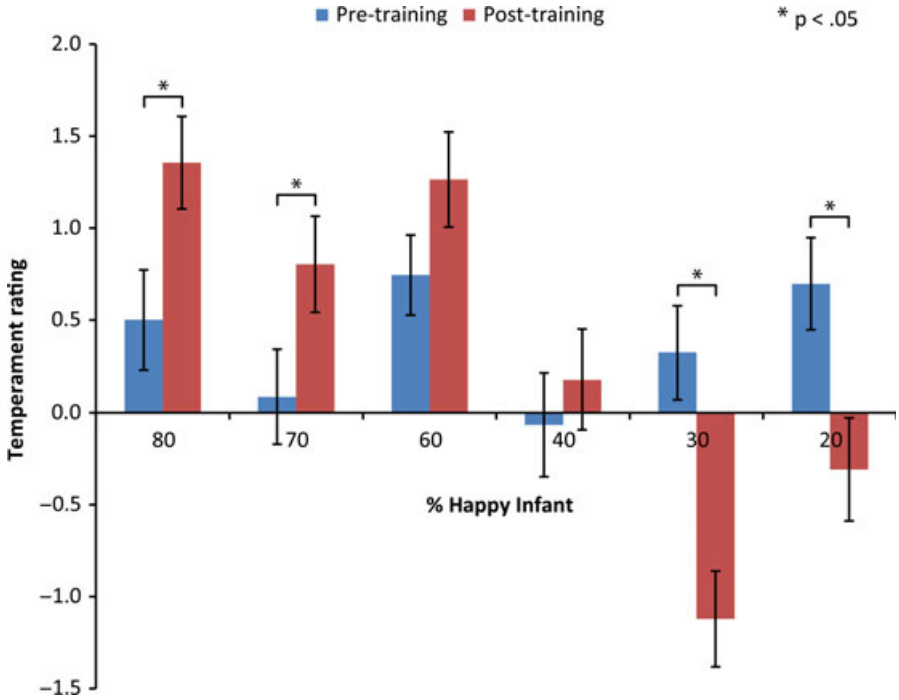

Figure 5 *Temperament ratings. Following training, the two happiest infants were rated as having significantly more positive temperament. The two saddest infants were rated as having significantly more negative temperament.*

It was hypothesized that the relationship between change in temperament ratings and change in number of key presses was mediated by change in cuteness ratings. To investigate this, two models of the relationship between temperament ratings and number of key presses were tested. In the first model, change in temperament ratings had a direct effect on change in number of key presses (R^2 = 0.11, p = .02). In the second model, two

variables were entered simultaneously: change in temperament and change in cuteness rating. In this model, the relationship between change in temperament ratings and change in number of key presses was mediated by changes in cuteness ratings ($R^2 = 0.21$, $p < .01$). Adding the mediating variable of 'change in cuteness rating' significantly improved the model fit ($\Delta R^2 = 0.10$, $p = .03$). The relationship between temperament and number of key presses decreased substantially when controlling for cuteness (model two, relationship between temperament and number of key presses: $R^2 = 0.02$, $p = .35$). The results of a Sobel test indicated that the association between temperament and key presses was significantly mediated by cuteness ($z = 2.50$, $p = .01$). This suggests that participants want to look at images of 'happier' infants for longer because they find them cuter.

Discussion

Our results demonstrate that infant 'temperament' can have a significant impact on adults' perception of cuteness and desire to view infant faces. We found that after only 6–8 minutes of learning about a set of previously unfamiliar infants, adults shifted their responses to each face based upon its experimentally established temperament. Infants who laughed more often than not were perceived as cuter and adults expended greater effort to view them after training. Infants who cried more often than not were not perceived as cuter and adults did not expend greater effort to view them after training.

Our learning task also had an effect on the extent to which adults perceived the infants as 'easy' or 'difficult', key descriptors of temperament. Infants who laughed more often were perceived as easier after training, whereas infants who cried more often were perceived as more difficult. These ratings indicate that participants were extrapolating what they had learned about the infant's mood (happy or sad) to ease of care. This link between infant mood and ease of care makes intuitive sense and is consistent with general conceptualizations of temperament (e.g. Van IJzendoorn & Bakermans-Kranenburg, 2012).

It is well established that physically attractive adults are perceived as having various positive personality qualities (e.g. Eagly *et al.*, 1991; Feingold, 1992; Langlois *et al.*, 2000). Here, we demonstrate the reverse relationship: personality, or its infant equivalent, 'temperament', can have an impact on both the perception of and motivation to view infants. To our knowledge, this is the first demonstration that infant temperament can alter cuteness perception. This provides a new way of thinking about the mechanisms through which infant temperament can impact on caregiving.

Previous studies of the perception of infant cuteness have examined factors related to the viewer, such as the adult's hormonal status, age and gender (Lobmaier, Sprengelmeyer, Wiffen & Perrett, 2010; Parsons *et al.*, 2011a; Sprengelmeyer *et al.*, 2013). Our results demonstrate that non-physical infant factors can also have an impact on cuteness perception. The findings reported here come from a within-subjects design, with a relatively homogenous sample of young females, at roughly the same stage of their hormonal cycle. Taken with previous findings, these results suggest that both viewer and infant nonphysical characteristics can alter perceptions of infant cuteness.

Strengths

A major issue in understanding the impact of infant temperament on early parenting responses is related to its measurement. A strength of our study is the use of a novel, experimental method to establish a sense of infant temperament. This method allows us to circumnavigate some of the issues inherent in the measurement of infant temperament, such as mood of the reporter. In addition, this method allows us to make inferences about the directionality of the relationship between infant temperament and perception of infant appearance. This is important because it may be that 'cuter' infants are perceived as having a more positive temperament, but not vice versa. Our results provide evidence of bidirectional influence. Cuter infants are perceived as having a more positive temperament and changes in temperament for the better result in the infant being perceived as being cuter.

The effect of infant cuteness on perception of temperament was small, and we did not manipulate cuteness. Nonetheless, our findings are in line with other work demonstrating an impact of cuteness on perception of temperament (e.g. Langlois *et al.*, 1995). The current paradigm would allow additional exploration of this relationship, for instance by deliberately using faces varying in cuteness.

Limitations

Our sample consisted of a relatively young sample of women, none of whom had children. We chose to test women only because previous studies of gender differences in processing of infant faces have reported mixed results (e.g. Hahn *et al.*, 2013; Parsons *et al.*, 2011a; Sprengelmeyer *et al.*, 2009) and the clearest effects have emerged from women. Related to this, there have

been some studies suggesting a greater response to infant stimuli in mothers compared to nonmothers (e.g. Giardino, Gonzalez, Steiner & Fleming, 2008; Nishitani, Doi, Koyama & Shinohara, 2011; Stallings, Fleming, Corter, Worthman & Steiner, 2001). Therefore, an important follow-up to this study would be to test a sample of mothers and indeed fathers, or adults with experience of caregiving in a professional context. It would also be of interest to examine whether our effects would extend to children's faces, given research suggesting adults' preference for younger compared with older children's faces (Luo, Li & Lee, 2011). Finally, the infant faces used here were Caucasian, and replication of our effects with infant faces of other ethnicities is clearly warranted.

Our manipulation of infant 'temperament' was necessarily simplified. We chose to focus on the intuitive, broad descriptors 'happy' and 'sad' in training participants to learn about the infants. Infant temperament is of course more complex than a happy or sad division, and can be defined in terms of infant activity level, fear, distress to limitations, soothability and so on (Rothbart, 1986). Our 'happy/sad' training, which led to changes in perceived infant 'easiness/difficulty', maps closely onto the temperament dimensions of 'emotionality', and 'irritability' or 'difficultness' (Bates, Freeland & Lounsbury, 1979; Deater-Deckard & Wang, 2012) rather than other important dimensions such as 'behavioural inhibition' (Kagan, 2012). Nevertheless, training adults to recognize infants as 'happy' or 'sad' is an intuitive starting point for understanding the relationship between temperament and adults' responses to infant faces. It would appear that temperament has clear consequences for how adults perceive 'bonnie' babies. Perception of cuteness is not based on physical features alone, and is modifiable through training.

References

Aharon, I., Etcoff, N., Ariely, D., Chabris, C.F., O'Connor, E., & Breiter, H.C. (2001). Beautiful faces have variable reward value: fMRI and behavioral evidence. *Neuron*, **32** (3), 537–551.

Bates, J.E., Freeland, C.A., & Lounsbury, M.L. (1979). Measurement of infant difficultness. *Child development*, **50** (3), 794–803.

Belsky, J., Bakermans-Kranenburg, M.J., & Van Ijzendoorn, M.H. (2007). For better and for worse: differential susceptibility to environmental influences. *Current Directions in Psychological Science*, **16** (6), 300–304.

Berkowitz, L., & Frodi, A. (1979). Reactions to a child's mistakes as affected by her/his looks and speech. *Social Psychology Quarterly*, **42** (4), 420–425.

Deater-Deckard, K., & Wang, Z. (2012). Anger and Irritability. In M. Zentner & R.L. Shiner (Eds.), *Handbook of temperament* (pp. 124–144). New York: Guilford.

deVries, M.W. (1984). Temperament and infant mortality among the Masai of East Africa. *American Journal of Psychiatry*, **141** (10), 1189–1194.

Dion, K.K. (1972). Physical attractiveness and evaluation of children's transgressions. *Journal of Personality and Social Psychology*, **24** (2), 207–213.

Eagly, A.H., Makhijani, M.G., Ashmore, R.D., & Longo, L.C. (1991). What is beautiful is good, but. .: a meta-analytic review of research on the physical attractiveness stereotype. *Psychological Bulletin*, **110** (1), 109–128.

Ellis, B.J., Boyce, W.T., Belsky, J., Bakermans-Kranenburg, M.J., & Van Ijzendoorn, M.H. (2011). Differential susceptibility to the environment: an evolutionary-neurodevelopmental theory. *Development and Psychopathology*, **23** (1), 7–28.

Feingold, A. (1992). Good-looking people are not what we think. *Psychological Bulletin*, **111** (2), 304–341.

Field, T.M., & Vega-Lahr, N. (1984). Early interactions between infants with cranio-facial anomalies and their mothers. *Infant Behavior and Development*, **7** (4), 527–530.

Frank, M.J., Seeberger, L.C., & O'Reilly, R.C. (2004). By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science*, **306** (5703), 1940–1943.

Frodi, A.M., Lamb, M.E., Leavitt, L.A., & Donovan, W.L. (1978). Fathers' and mothers' responses to infant smiles and cries. *Infant Behavior and Development*, **1** (2), 187–198.

Giardino, J., Gonzalez, A., Steiner, M., & Fleming, A.S. (2008). Effects of motherhood on physiological and subjective responses to infant cries in teenage mothers: a comparison with non-mothers and adult mothers. *Hormones and Behavior*, **53** (1), 149–158.

Glocker, M.L., Langleben, D.D., Ruparel, K., Loughead, J.W., Valdez, J.N., Griffin, M.D., Sachser, N., & Gur, R.C. (2009). Baby schema modulates the brain reward system in nulliparous women. *Proceedings of the National Academy of Sciences of the United States of America*, **106** (22), 9115–9119.

Hahn, A.C., Xiao, D., Sprengelmeyer, R., & Perrett, D.I. (2013). Gender differences in the incentive salience of adult and infant faces. *Quaterly Journal of Experimental Psychology*, **66** (1), 200–208.

Kagan, J. (2007). A trio of concerns. *Perspectives on Psychological Science*, **2** (4), 361–373.

Kagan, J. (2009). Two is better than one. *Perspectives on Psychological Science*, **4** (1), 22–23.

Kagan, J. (2012). The biography of behavioural inhibition In M. Zentner & R.L. Shiner (Eds.), *Handbook of temperament* (pp. 69–82). New York: Guilford Press.

Karraker, K.H., & Stern, M. (1990). Infant physical attractiveness and facial expression: effects on adult perceptions. *Basic Applied Social Psychology*, **11**, 371–385.

Kniffin, K.M., & Wilson, D.S. (2004). The effect of nonphysical traits on the perception of physical attractiveness: three

naturalistic studies. *Evolution and Human Behavior*, **25** (2), 88–101.

Kringelbach, M.L., Lehtonen, A., Squire, S., Harvey, A.G., Craske, M.G., Holliday, I.E., Green, A.L., Aziz, T.Z., Hansen, P.C., Cornelissen, P.L., & Stein, A. (2008). A specific and rapid neural signature for parental instinct. *PLoS ONE*, **3** (2), e1664.

Kringelbach, M.L., & Rolls, E.T. (2003). Neural correlates of rapid reversal learning in a simple model of human social interaction. *NeuroImage*, **20** (2), 1371–1383.

Langlois, J.H., Kalakanis, L., Rubenstein, A.J., Larson, A., Hallam, M., & Smoot, M. (2000). Maxims or myths of beauty? A meta-analytic and theoretical review. *Psychological Bulletin*, **126** (3), 390–414.

Langlois, J.H., Ritter, J.M., Casey, R.J., & Sawin, D.B. (1995). Infant attractiveness predicts maternal behaviors and attitudes. *Developmental Psychology*, **31** (3), 464–472.

Lobmaier, J.S., Sprengelmeyer, R., Wiffen, B., & Perrett, D.I. (2010). Female and male responses to cuteness, age and emotion in infant faces. *Evolution and Human Behavior*, **31** (1), 16–21.

Lorenz, K. (1943). Die angeborenen Formen moeglicher Erfahrung. *Z Tierpsychol*, **5**, 235–409.

Lorenz, K. (1971). *Studies in animal and human behavior*, Vol. 2. London: Methuen.

Luo, L.Z., Li, H., & Lee, K. (2011). Are children's faces really more appealing than those of adults? Testing the baby schema hypothesis beyond infancy. *Journal of Experimental Child Psychology*, **110** (1), 115–124.

Maier ., R.A. Jr, Holmes, D.L., Slaymaker, F.L., & Reich, J.N. (1984). The perceived attractiveness of preterm infants. *Infant Behavior and Development*, **7** (4), 403–414.

Mäntymaa, M., Puura, K., Luoma, I., Salmelin, R.K., & Tamminen, T. (2006). Mother's early perception of her infant's difficult temperament, parenting stress and early mother–infant interaction. *Nordic Journal of Psychiatry*, **60** (5), 379–386.

Murray, L., Hentges, F., Hill, J., Karpf, J., Mistry, B., Kreutz, M., Woodall, P., Moss, T., & Goodacre, T. (2008). The effect of cleft lip and palate, and the timing of lip repair on mother–infant interactions and infant development. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, **49** (2), 115–123.

Nishitani, S., Doi, H., Koyama, A., & Shinohara, K. (2011). Differential prefrontal response to infant facial emotions in mothers compared with non-mothers. *Neuroscience Research*, **70** (2), 183–188.

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D.M., & Dolan, R.J. (2003). Beauty in a smile: the role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, **41** (2), 147–155.

Parks, F.R., & Kennedy, J.H. (2007). The impact of race, physical attractiveness, and gender on education majors' and teachers' perceptions of student competence. *Journal of Black Studies*, **37** (6), 936–943.

Parsons, C.E., Young, K.S., Kumari, N., Stein, A., & Kringelbach, M.L. (2011a). The motivational salience of infant faces is similar for men and women. *PLoS ONE*, **6** (5), e20632.

Parsons, C.E., Young, K.S., Mohseni, H., Woolrich, M.W., Thomsen, K.R., Joensson, M., Murray, L., Goodacre, T., Stein, A., & Kringelbach, M.L. (2013a). Minor structural abnormalities in the infant face disrupt neural processing: a unique window into early caregiving responses. *Social Neuroscience*, **8** (4), 268–274. doi: 10.1080/17470919.2013.795189

Parsons, C.E., Young, K.S., Murray, L., Stein, A., & Kringelbach, M.L. (2010). The functional neuroanatomy of the evolving parent–infant relationship. *Progress in Neurobiology*, **91** (3), 220–241.

Parsons, C.E., Young, K.S., Parsons, E., Dean, A., Murray, L., Goodacre, T., Dalton, L., Stein, A., & Kringelbach, M.L. (2011b). The effect of cleft lip on adults' responses to faces: cross-species findings. *PLoS ONE*, **6** (10), e25897.

Parsons, C.E., Young, K.S., Parsons, E., Stein, A., & Kringelbach, M.L. (2012). Listening to infant distress vocalizations enhances effortful motor performance. *Acta Paediatrica, International Journal of Paediatrics*, **101** (4), e189–e191.

Parsons, C.E., Young, K.S., Joensson, M., Brattico, E., Hyam, J.A., Stein, A., Green, A.L., Aziz, T.Z., & Kringelbach, M.L. (2013b). Ready for action: a role for the human midbrain in responding to infant vocalisations *Social Cognitive and Affective Neuroscience*.

Paulussen-Hoogeboom, M.C., Stams, G.J.J.M., Hermanns, J.M.A., & Peetsma, T.T.D. (2007). Child negative emotionality and parenting from infancy to preschool: a meta-analytic review. *Developmental Psychology*, **43** (2), 438–453.

Ritter, J.M., Casey, R.J., & Langlois, J.H. (1991). Adults' responses to infants varying in appearance of age and attractiveness. *Child Development*, **62** (1), 68–82.

Ritts, V., Patterson, M.L., & Tubbs, M.E. (1992). Expectations, impressions, and judgments of physically attractive students: a review. *Review of Educational Research*, **62** (4), 413–426.

Rothbart, M.K. (1986). Longitudinal observation of infant temperament. *Developmental Psychology*, **22** (3), 356–365.

Rothbart, M.K., & Bates, J.E. (2006). Temperament. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Social, emotional and personality development* (5th edn., pp. 105–176). New York: Wiley.

Said, C.P., Haxby, J.V., & Todorov, A. (2011). Brain systems for assessing the affective value of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, **366** (1571), 1660–1670.

Seifer, R., Sameroff, A.J., Barrett, L.C., & Krafchuk, E. (1994). Infant temperament measured by multiple observations and mother report. *Child Development*, **65** (5), 1478–1490.

Sprengelmeyer, R., Lewis, J., Hahn, A., & Perrett, D.I. (2013). Aesthetic and incentive salience of cute infant faces: studies of observer sex, oral contraception and menstrual cycle. *PLoS ONE*, **8** (5), e65844.

Sprengelmeyer, R., Perrett, D.I., Fagan, E.C., Cornwell, R.E., Lobmaier, J.S., Sprengelmeyer, A., Aasheim, H.B.M., Black, I.M., Cameron, L.M., Crow, S., Milne, N., Rhodes, E.C., & Young, A.W. (2009). The cutest little baby face: a hormonal link to sensitivity to cuteness in infant faces. *Psychological Science*, **20** (2), 149–154.

Stallings, J., Fleming, A.S., Corter, C., Worthman, C., & Steiner, M. (2001). The effects of infant cries and odors on sympathy, cortisol, and autonomic responses in new mothers and nonpostpartum women. *Parenting: Science and Practice*, **1**, 77–100.

Stephan, C.W., & Langlois, J.H. (1984). Baby beautiful: adult attributions of infant competence as a function of infant attractiveness. *Child Development*, **55** (2), 576–585.

Swami, V., Furnham, A., Chamorro-Premuzic, T., Akbar, K., Gordon, N., Harris, T., Finch, J., & Tovée, M.J. (2010). More than just skin deep? Personality information influences men's ratings of the attractiveness of women's body sizes. *Journal of Social Psychology*, **150** (6), 628–647.

Tinbergen, N. (1951). *The study of instinct*. Oxford: Clarendon Press.

Van IJzendoorn, M.H., & Bakermans-Kranenburg, M.J. (2012). Integrating temperament and attachment: the differential susceptibility paradigm. In M. Zentner & R.L. Shiner (Eds.), *Handbook of temperament* (pp. 403–419). New York: Guilford Press.

Volk, A., & Quinsey, V.L. (2002). The influence of infant facial cues on adoption preferences. *Human Nature*, **13** (4), 437–455.

Waller, K.L., Volk, A., & Quinsey, V.L. (2004). The effect of infant fetal alcohol syndrome facial features on adoption preference. *Human Nature*, **15** (1), 101–117.

Wolke, D., Rizzo, P., & Woods, S. (2002). Persistent infant crying and hyperactivity problems in middle childhood. *Pediatrics*, **109** (6), 1054–1060.

Yamamoto, R., Ariely, D., Chi, W., Langleben, D.D., & Elman, I. (2009). Gender differences in the motivational processing of babies are determined by their facial attractiveness. *PLoS ONE*, **4** (6), e6042.

Young, K.S., Parsons, C.E., Stein, A., & Kringelbach, M.L. (2012). Interpreting infant vocal distress: the ameliorative effect of musical training in depression. *Emotion*, **12** (6), 1200–1205.

Youngstrom, E., Izard, C., & Ackerman, B. (1999). Dysphoria-related bias in maternal ratings of children. *Journal of Consulting and Clinical Psychology*, **67** (6), 905–916.

Zentner, M., & Bates, J.E. (2008). Child temperament: an integrative review of concepts, research programs, and measures. *European Journal of Developmental Science*, **2**, 7–37.

Received: 29 November 2012
Accepted: 11 July 2013